Metagenomics:

Challenges and Opportunities

for

Microbial Control

Karen E. Nelson
Director
Maryland

J. Craig Venter™
INSTITUTE

J. Craig Venter Institute, Rockville, MD

Formerly The Institute for Genomic Research (TIGR)

J. Craig Venter™
INSTITUTE

J. Craig Venter Institute

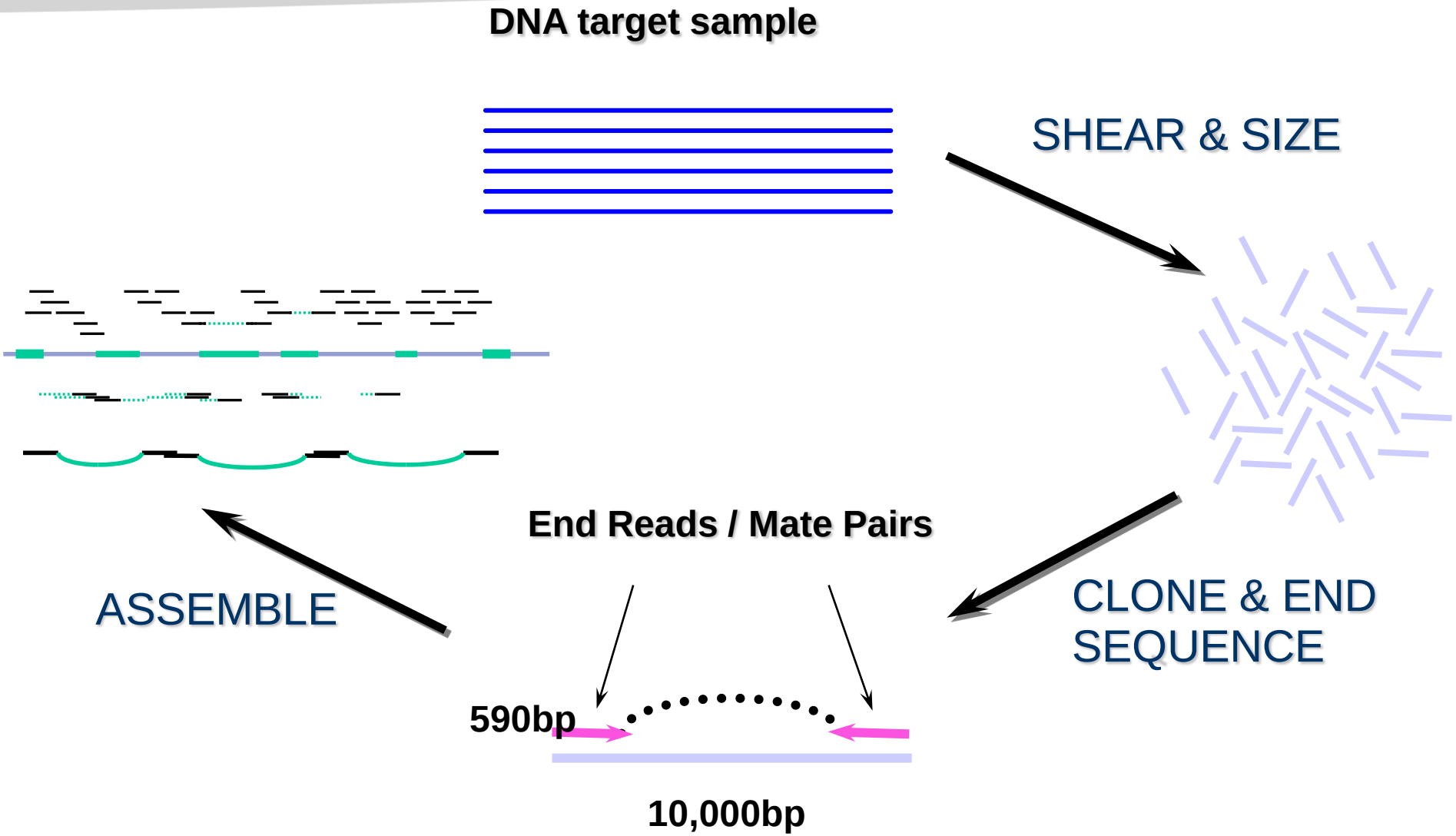
~ 350 staff & scientists

Rockville, MD and La Jolla, CA

- **Microbial, Plant, Environmental, Human & Evolutionary biology**
- **Genetics and Genomics**
- **High-throughput DNA sequencing**
- **Functional Genomics**
- **Bioinformatics**
- **Information technology**
- **Genomic policy research**
- **Environmental policy research**

J. Craig Venter™ INSTITUTE

Developed the Whole Genome Shotgun Sequencing (WGS) approaches



J. Craig Venter INSTITUTE

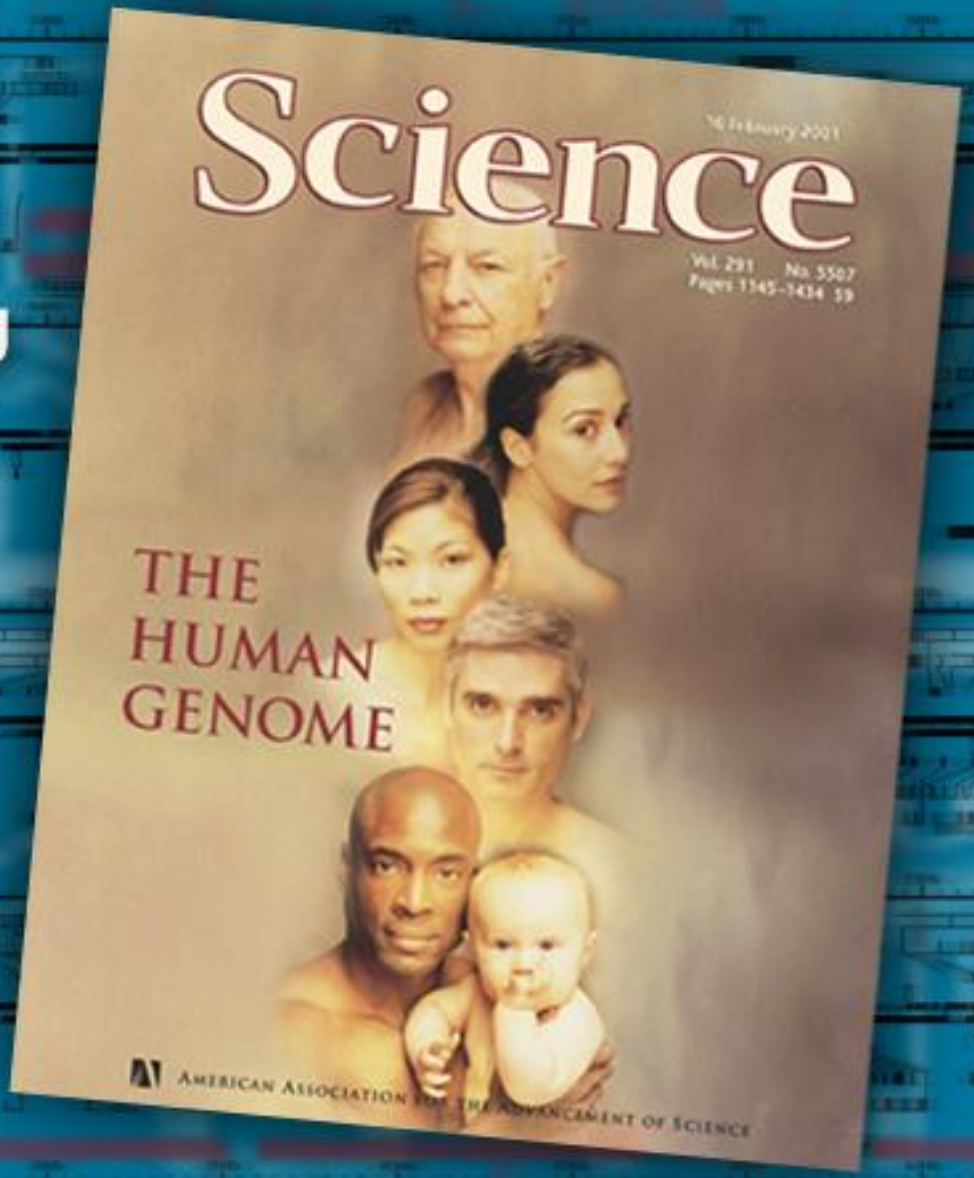
First Complete Genome of a Free Living Organism
Haemophilus influenzae

J. Craig Venter™ INSTITUTE

First Human Genome

CELERA

Sequencing the Human Genome

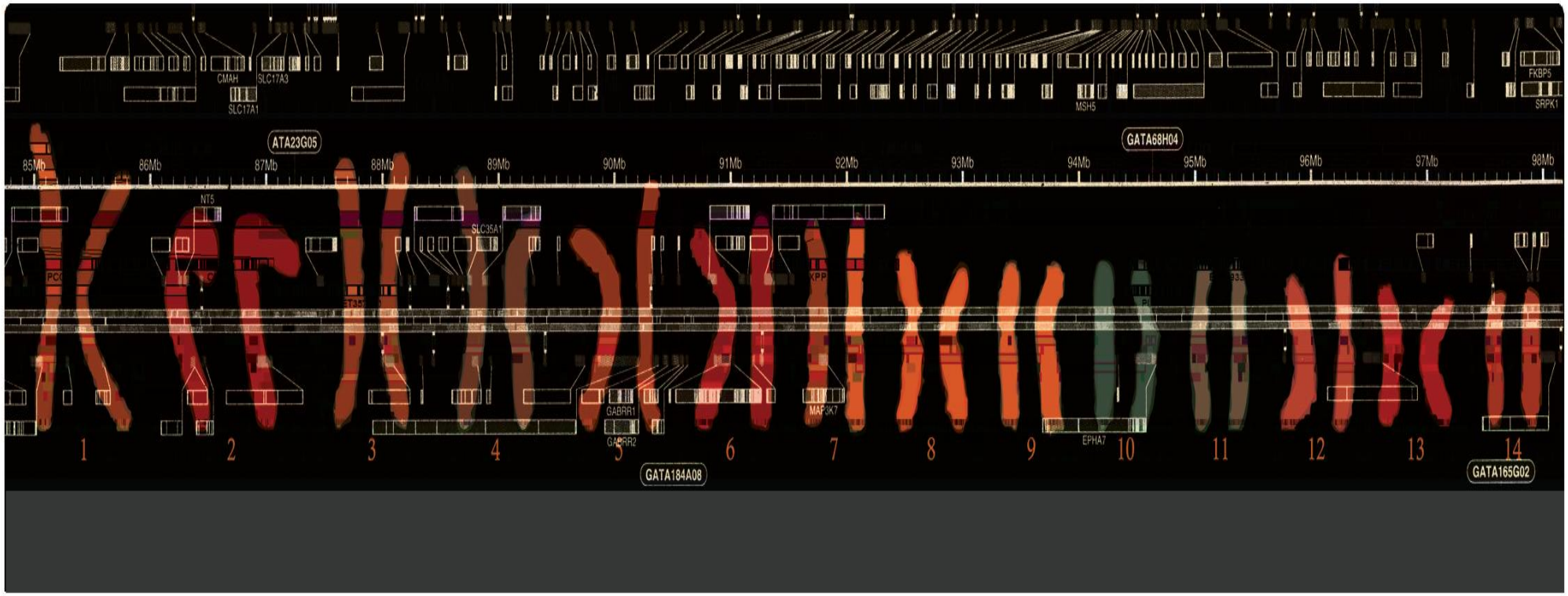


First Diploid Human Genome

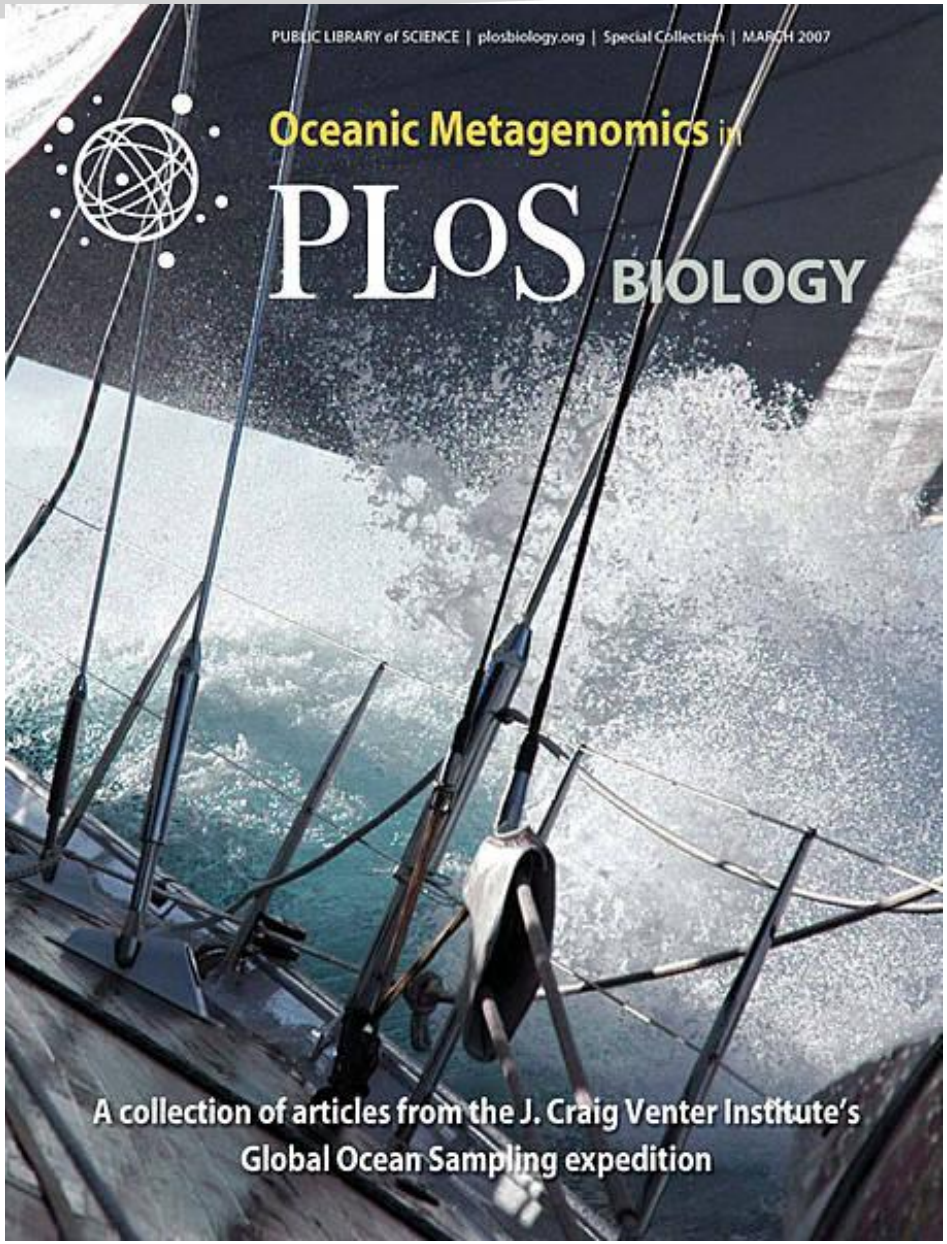
PLoS BIOLOGY
a peer-reviewed open-access journal published by the Public Library of Science

RESEARCH ARTICLE OPEN ACCESS

The Diploid Genome Sequence of an Individual Human

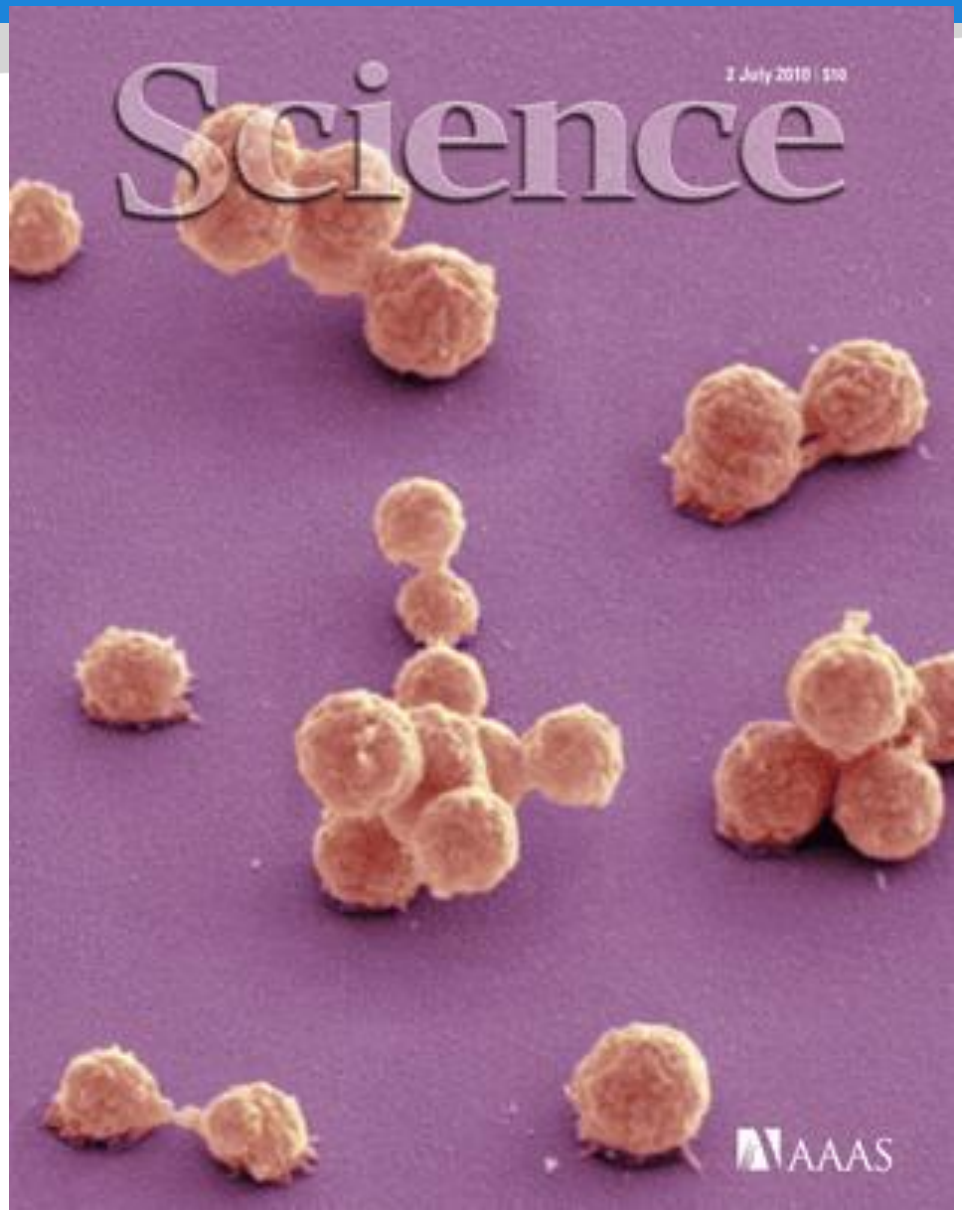


Metagenomic Sequencing and Analysis of Environmental and Human Microbiome



J. Craig Venter™
INSTITUTE

First Synthetic Self Replicating Bacterial Cell



J. Craig Venter™
INSTITUTE

Also…

- *Haemophilus influenzae (1995)*
- Reverse vaccinology (2000)
- Human microbiome (2006)
- Diploid human genome (2007)
- Genome transplantation (2007)
- Global Ocean Survey (2007)
- Synthetic microbial genome (2008)
- >7,000 influenza genomes (ongoing)
- Sequenced most major pathogens (e.g. TB, malaria, cholera, *T. parva*, *T. cruzi*)

J. Craig Venter™ INSTITUTE

Genomics Applications

Surveillance of Emerging Infectious Diseases:

- Diseases targeted whole genome sequencing
- Diseases targeted variation discovery, analysis and monitoring

Human Genomic Medicine:

- Whole genome sequencing
- Variation discovery and analysis
- Expression profiling
- Functional tag analysis
- Methylation analysis
- Genomic based diagnostics

Human and Environmental Metagenomic Discovery, Analysis and Monitoring:

- Microbial, fungal and viral population profiling, analysis and monitoring

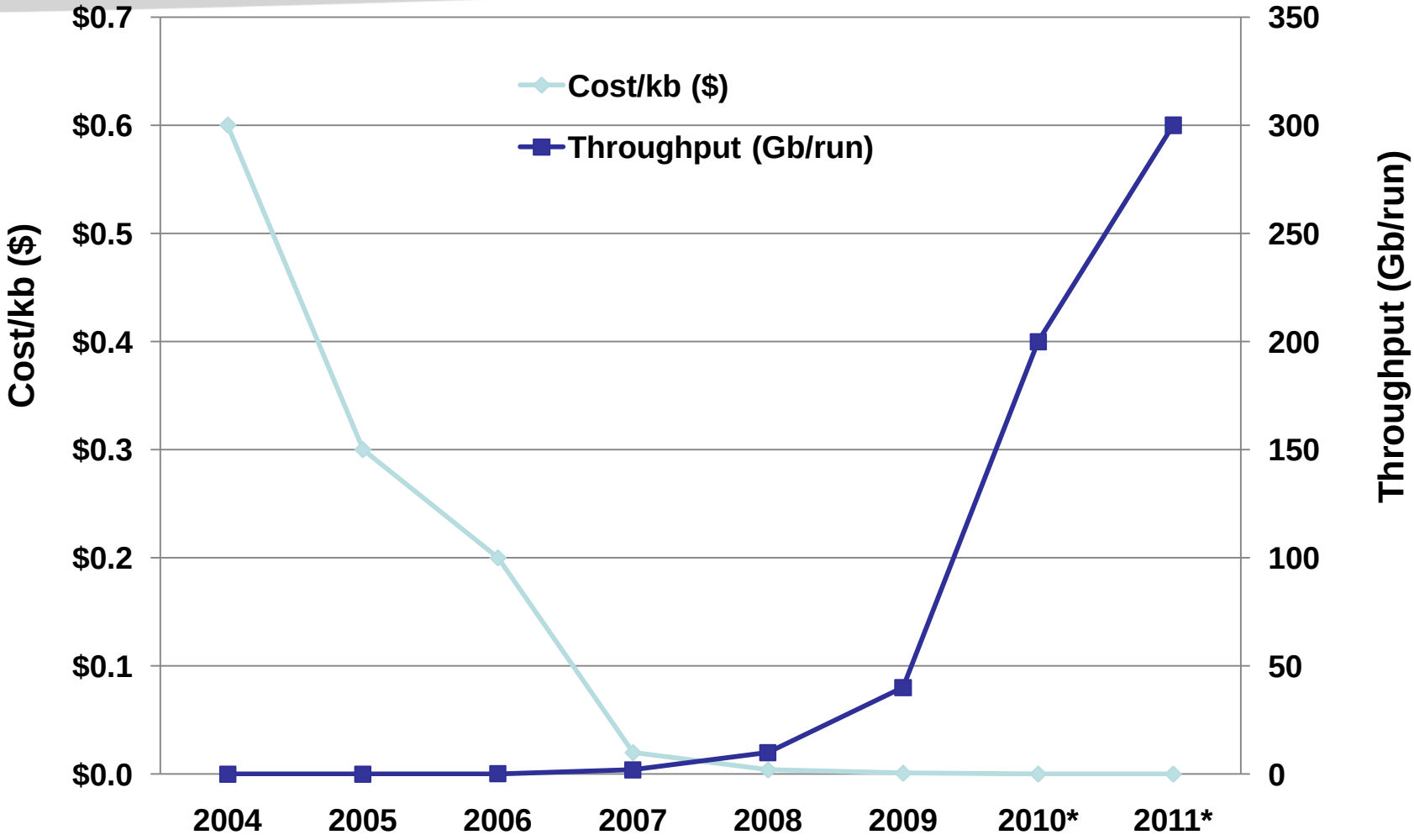
J. Craig Venter™ INSTITUTE

Sequencing Technologies Use at JCVI

	Read length bp	Throughput /machine	Run time	Throughput /day	Accuracy	Cost/Gbp
Sanger	600-800	75,000bp	30-60 min	1-2 Mb	> QV 30	$3,000,000
454	400-600	400 Mb	7 hr	800 Mb	QV 20	$30,000
Illumina	35-150	up to 200 Gb	3-12 days	16 Gb	~85% bases > QV30	$900
SOLiD	35-50	up to 100 Gb	3-12 days	8 Gb	~80% bases > QV30	$900

J. Craig Venter™ INSTITUTE

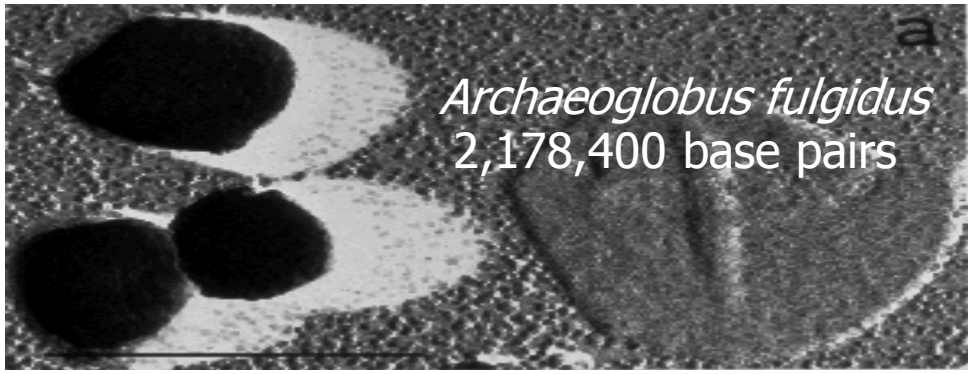
Sequencing Throughput vs Cost

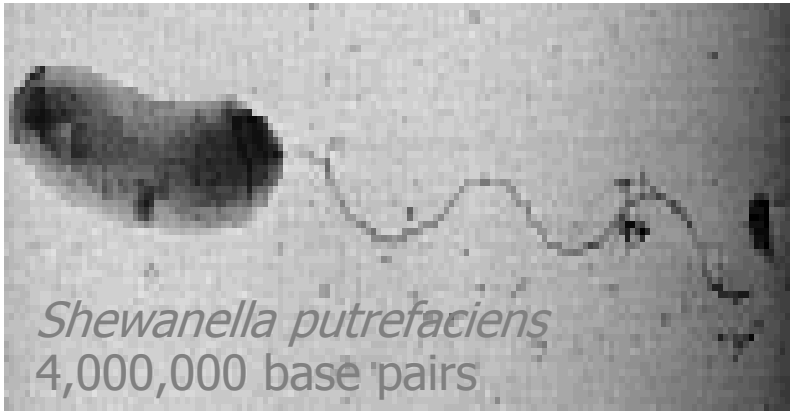


J. Craig Venter™ INSTITUTE

Microbial Genomics

- Disease-causing Organisms
- Environmentally Significant Organisms
- Understand novel metabolisms
- Identify potential genes and pathways for bioremediation
- Understand adaptation to extreme environments
- Development of novel informatics approaches

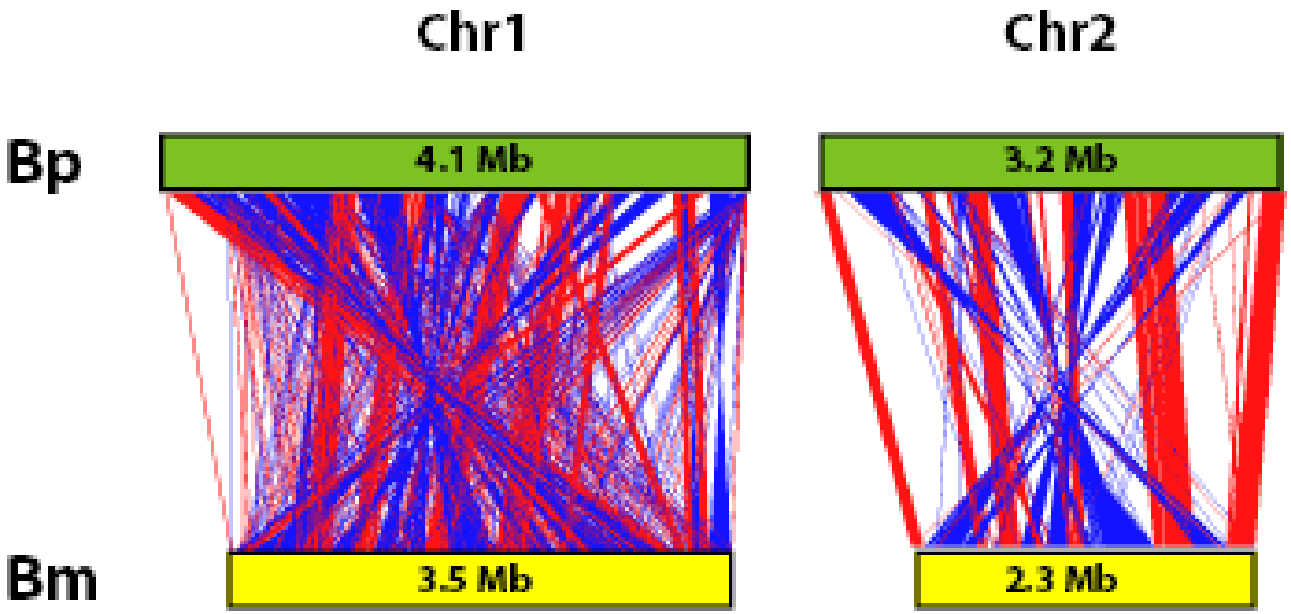




J. Craig Venter™
INSTITUTE

Burkholderia mallei and glanders

- **Disease of equines, described by Hippocrates and Aristotle**
- **Weaponized by Soviets; Used as a weapon Civil War, WWI, WWII**

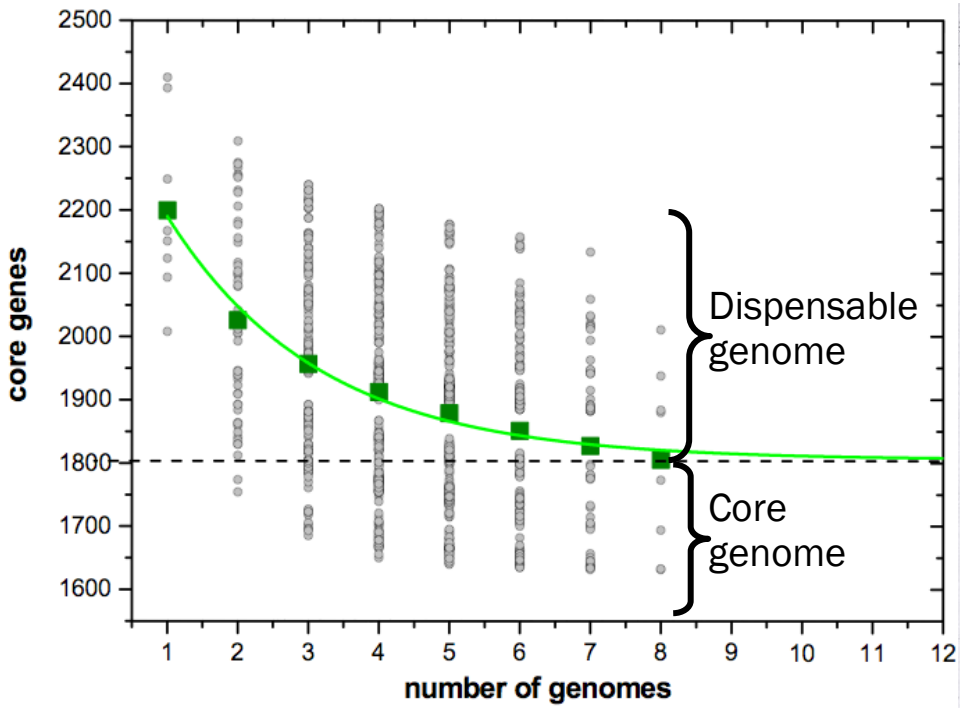


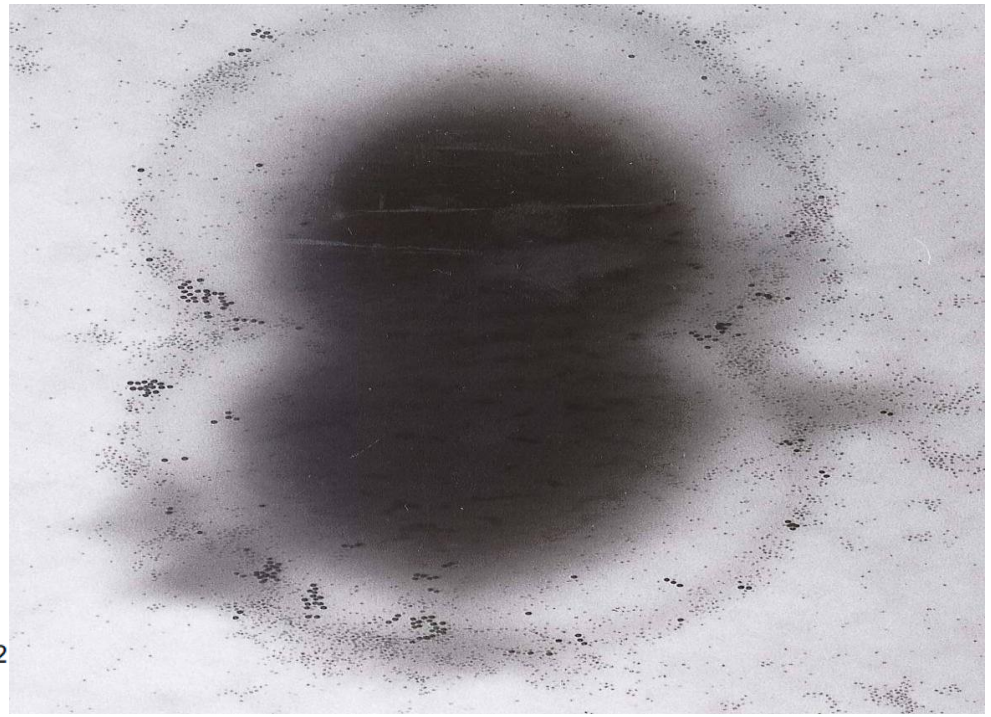
Species *B. pseudomallei* vs. *B. mallei*
soil and water *vs.* *obligate mammalian parasite*
melioidosis *vs.* *glanders*

Red and Blue: Opposite Orientations

Genomic diversity - Pan-Genome

The pan-genome concept - The entire gene pool accessible to the species is much larger than the genome of any individual strain or isolate





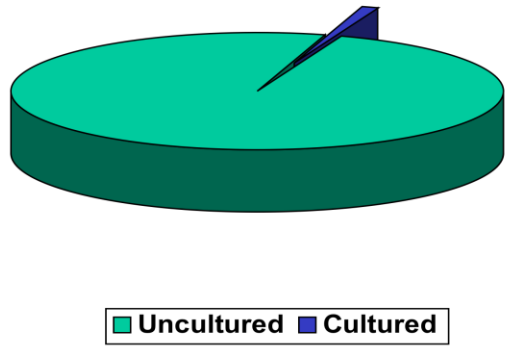
In the case of GBS, the pan-genome is unbounded and each new genome sequence is predicted to provide an average of 33 new genes not previously identified

J. Craig Venter™ INSTITUTE

Microbial Diversity and Metagenomics

For the most part, we have sequenced and analyzed genomes mainly of bacteria/archaea that are culturable.

- These culturables represent less than 1% of total microbial diversity.
- We know very little about the remaining 99% of microbial diversity.
- Traditional methods for measuring diversity inclusive of culturing and 16S PCR are known to have bias
- How do we study the 99% of uncultured (and unidentified) bacteria/archaea?



J. Craig Venter™ INSTITUTE

- Represents an ecosystem with relatively low microbial diversity
- Long history of physical & biological oceanography data

Sargasso Sea study

- Venter and colleagues at the JCVI
- Generated 1,987,936 DNA reads
- Approximately 1, 625 Mb of DNA
- 1.2 million new genes identified
- ~1,412 rRNA genes
- Estimated 1,800 species
- 12 complete genomes recovered
- Demonstration of the power of genomics

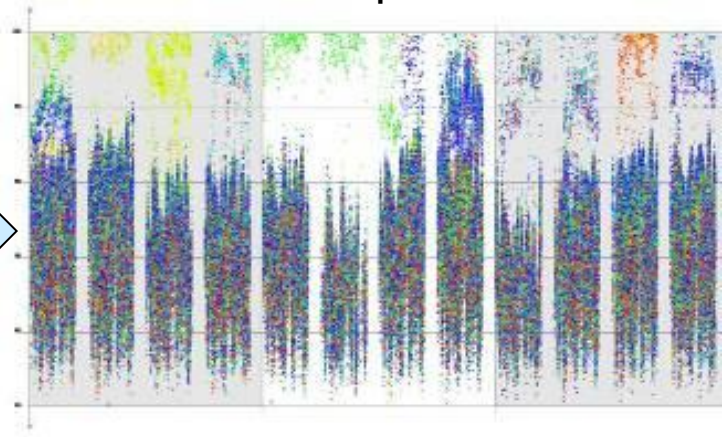
J. Craig Venter™ INSTITUTE

J. Craig Venter™
INSTITUTE

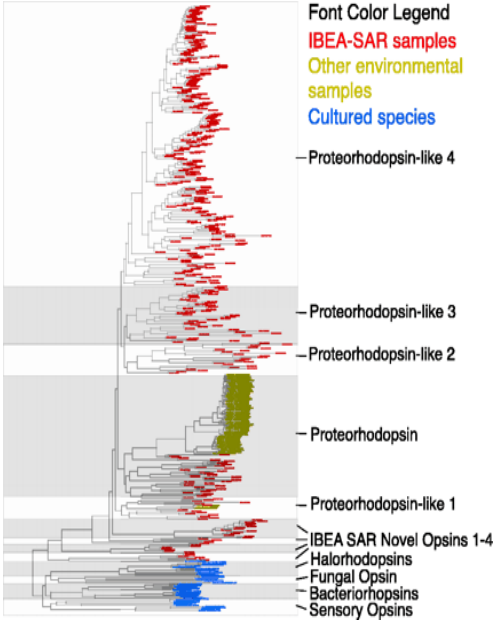
Global Ocean Sampling and Analysis

Sampling and Sequencing

Tool Development

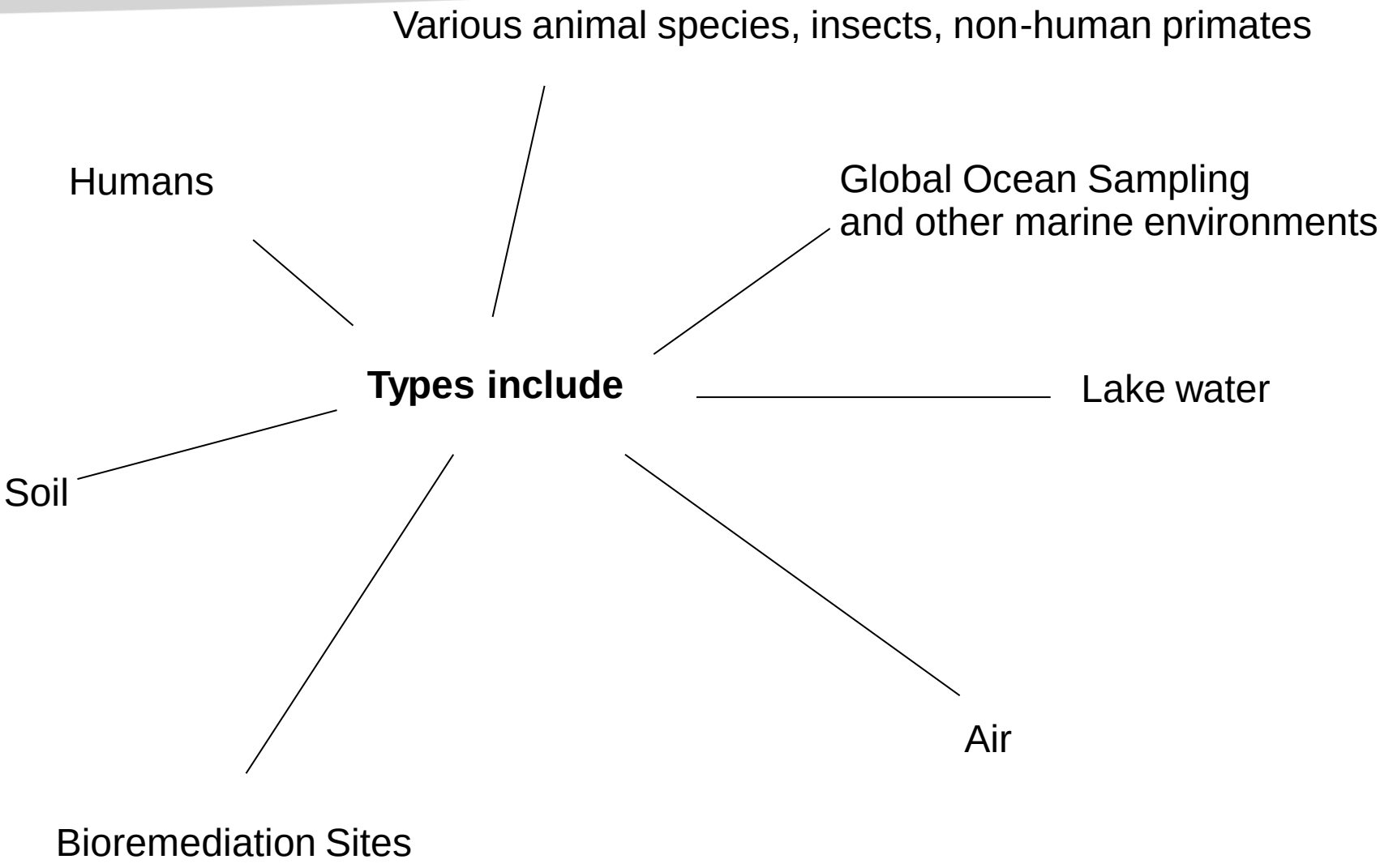


Data Analysis



J. Craig Venter™ INSTITUTE

Metagenomic projects



J. Craig Venter™
INSTITUTE

Human Microbiome
Metagenomics, Health and Disease

J. Craig Venter™
INSTITUTE

Human Microbiome

- **Collective of the human microbiome exceeds the number of human cells (somatic and germ cells) by at least an order of magnitude**

- **The majority of the human microbiome remains unknown**

- **Many of these microbial interactions endow or enhance human physiology including processes related to development, nutrition, immunity and resistance to pathogens**

- **Many relationships between the human host and microbiome remain to be determined**

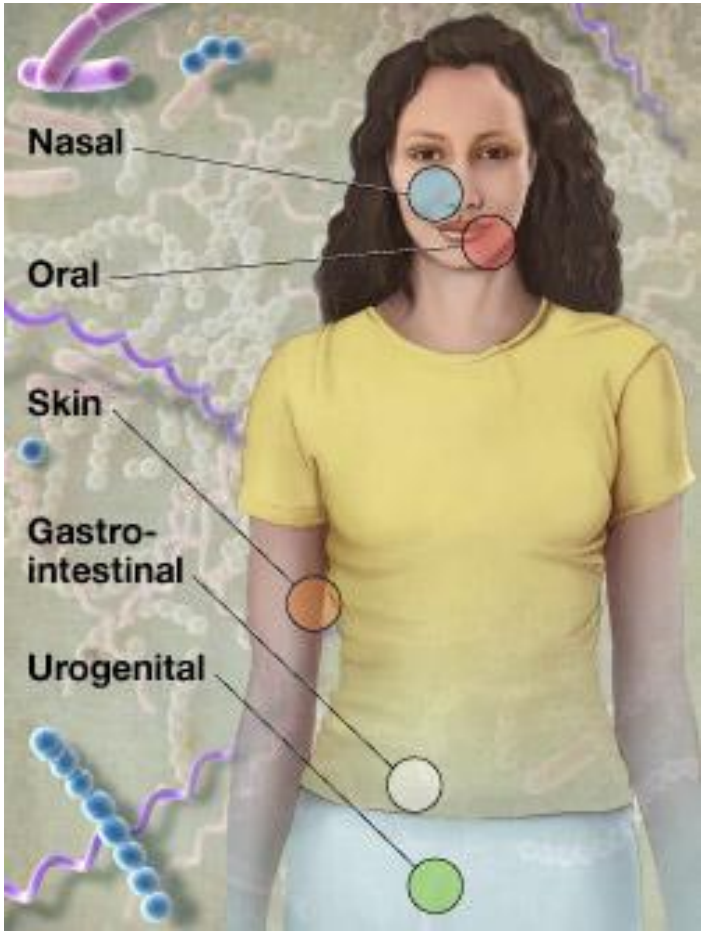
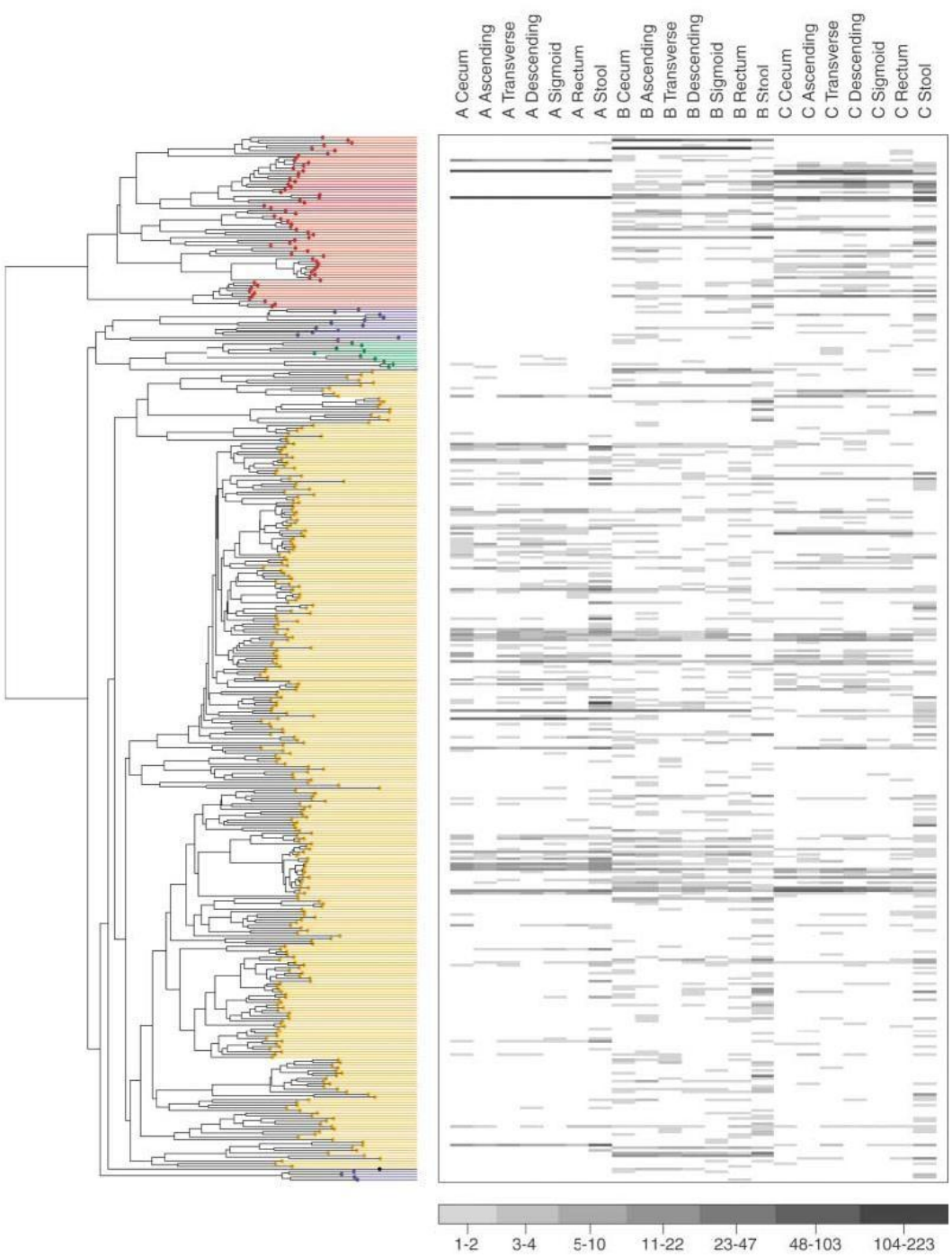


image courtesy of the NIH HMP website
http://nihroadmap.nih.gov/hmp/

J. Craig Venter™
INSTITUTE

Human Colon



Mucosal samples were obtained during colonoscopy from healthy-appearing sites within the six major subdivisions of the human colon:
cecum
ascending colon
transverse colon
descending colon
sigmoid colon
rectum.

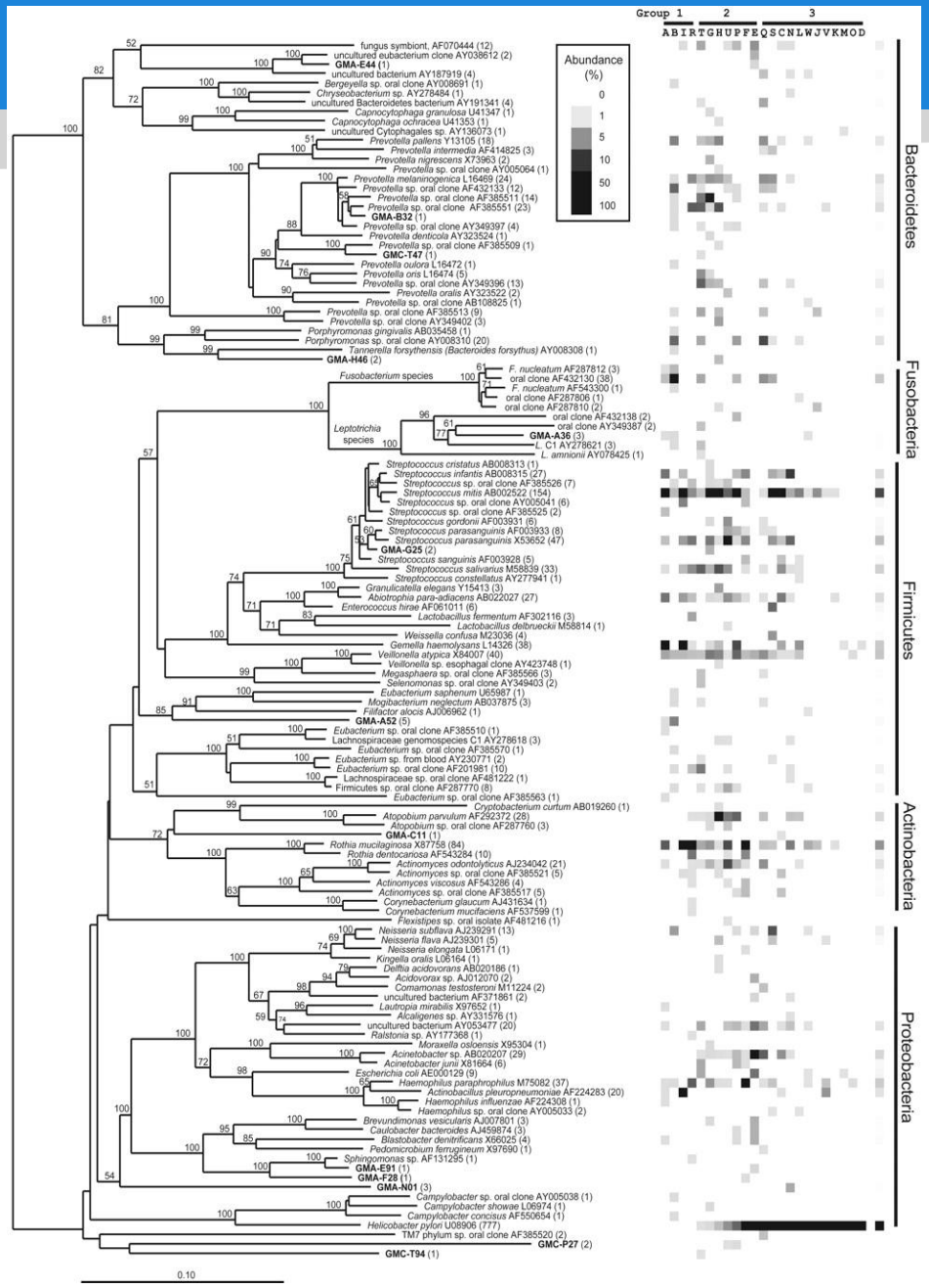
Fecal samples were collected from each subject 1 month following colonoscopy.

From 11,831 bacterial and 1524 archaeal 16S sequences, identified 395 phylotypes

Eckburg et at., 2005 Science 308(5728):1635-8.

J. Craig Venter™ INSTITUTE

Stomach



1,833 full-length 16S sequences

Described 128 16S rDNA phylotypes

Derived from 23 human subjects

Bik, E.M. et al. (2006) PNAS 103, 732-737

PNAS

- First human metagenomic paper
- Investigated the gastrointestinal tract (via fecal samples) of two healthy adults
- 78 Mbp
- 2062 amplified 16S rDNA

RESEARCH ARTICLE

Metagenomic Analysis of the Human Distal Gut Microbiome

Steven R. Gill,[1]*‡ Mihai Pop,[1]† Robert T. DeBoy,[1] Paul B. Eckburg,[2,3,4] Peter J. Turnbaugh,[5] Buck S. Samuel,[5] Jeffrey I. Gordon,[5] David A. Relman,[2,3,4] Claire M. Fraser-Liggett,[1,6] Karen E. Nelson[1]

The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. We analyzed ~78 million base pairs of unique DNA sequence and 2062 polymerase chain reaction–amplified 16S ribosomal DNA sequences obtained from the fecal DNAs of two healthy adults. Using metabolic function analyses of identified genes, we compared our human genome with the average content of previously sequenced microbial genomes. Our microbiome has significantly enriched metabolism of glycans, amino acids, and xenobiotics; methanogenesis; and 2-methyl-D-erythritol 4-phosphate pathway–mediated biosynthesis of vitamins and isoprenoids. Thus, humans are superorganisms whose metabolism represents an amalgamation of microbial and human attributes.

Our body surfaces are home to microbial communities whose aggregate [illegible] ≥100 times as many genes as our 2.85–billion base pair (bp) human genome (*1*). Therefore, a [illegible]

of single organisms, recent reports from Venter *et al.* (*9*) and Baker *et al.* (*10*) have demonstrated the utility of this approach for studying mixed microbial communities. Variations in the relative abundance of each member of the microbial community and their respective genome sizes determine the final depth of sequence coverage for any organism at a particular level of sequencing. This means that the genome sequences of abundant species will be well represented in a set of random shotgun reads, whereas lower abundance species may be represented by a small number of sequences. In fact, the size and depth of coverage (computed as the ratio between the total length of the reads placed into contigs and the total size of the contigs) of genome assemblies generated from a metagenomics project can provide information on relative species abundance.

A total of 65,059 and 74,462 high-quality sequence reads were generated from random DNA libraries created with fecal specimens of two healthy humans (subjects 7 and 8). These two subjects, ages 28 and 37, female and male, [illegible]

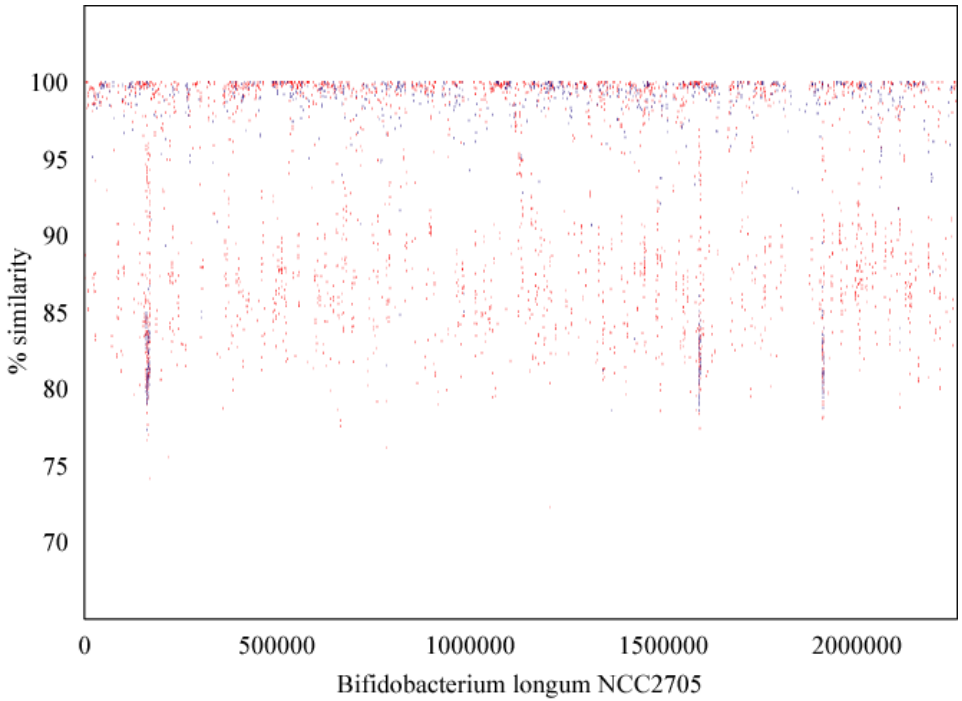
Gill et al, *Science* 2006

J. Craig Venter™ INSTITUTE

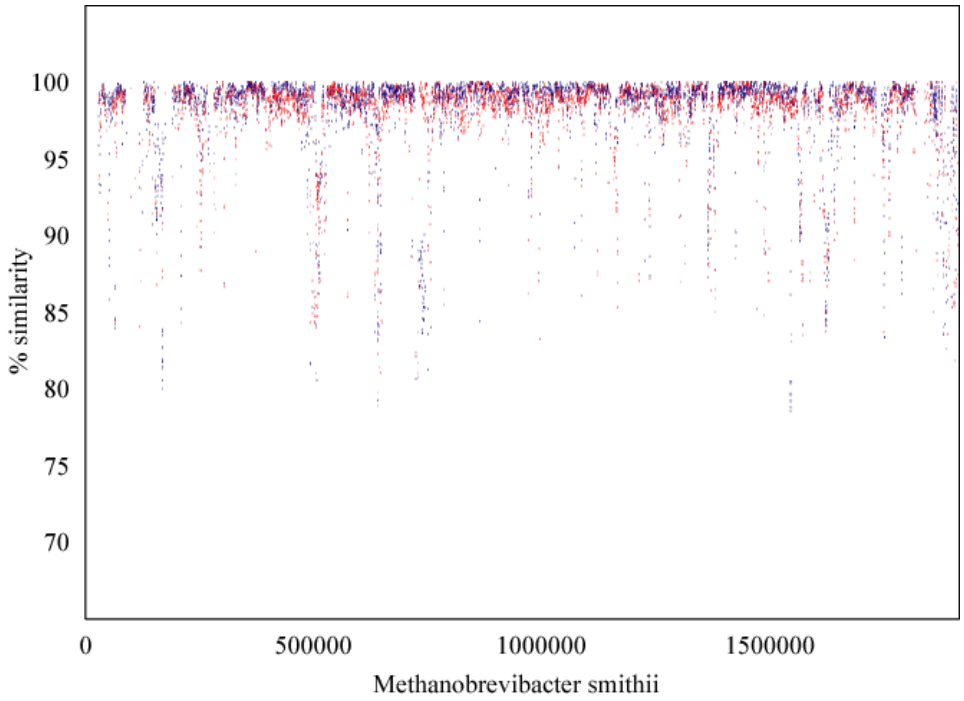
Genomics of the Human Colonic Microbiome

Genome diversity

Bifidobacterium longum



Methanobrevibacter smithii



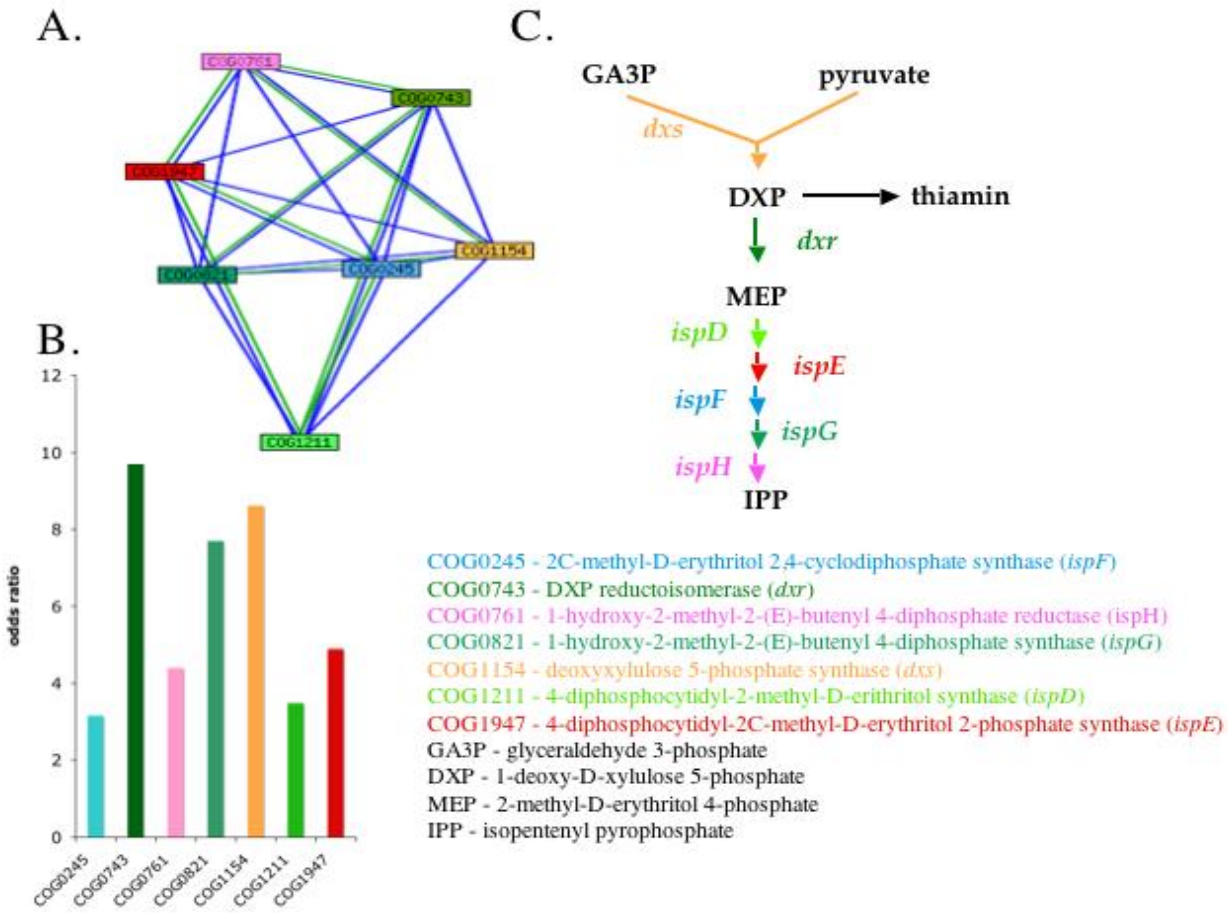
Gill et al, *Science* 2006

J. Craig Venter™ INSTITUTE

Genomics of the Human Colonic Microbiome

Perhaps the most interesting question is to delineate ways in which our microbiome endows us with physiological properties that we have not evolved on our own-----what is the metabolic potential of the colonic microbiota?

glycan metabolism
amino acid metabolism
xenobiotic metabolism
isoprenoid biosynthesis
vitamins



J. Craig Venter INSTITUTE

NIH Roadmap Human Microbiome Project

- **Budget $157 million 2007-2013**
- **Goal: Characterize the microbes that inhabit the human body and examine whether changes in the microbiome can be related to health and disease**
- **Feasibility project designed to determine the value of microbial metagenomics to biomedical research**
- **Community Resource Project-generate reagents and data sets; rapidly placed in public domain**
- **Continuous Scientific Community Input**

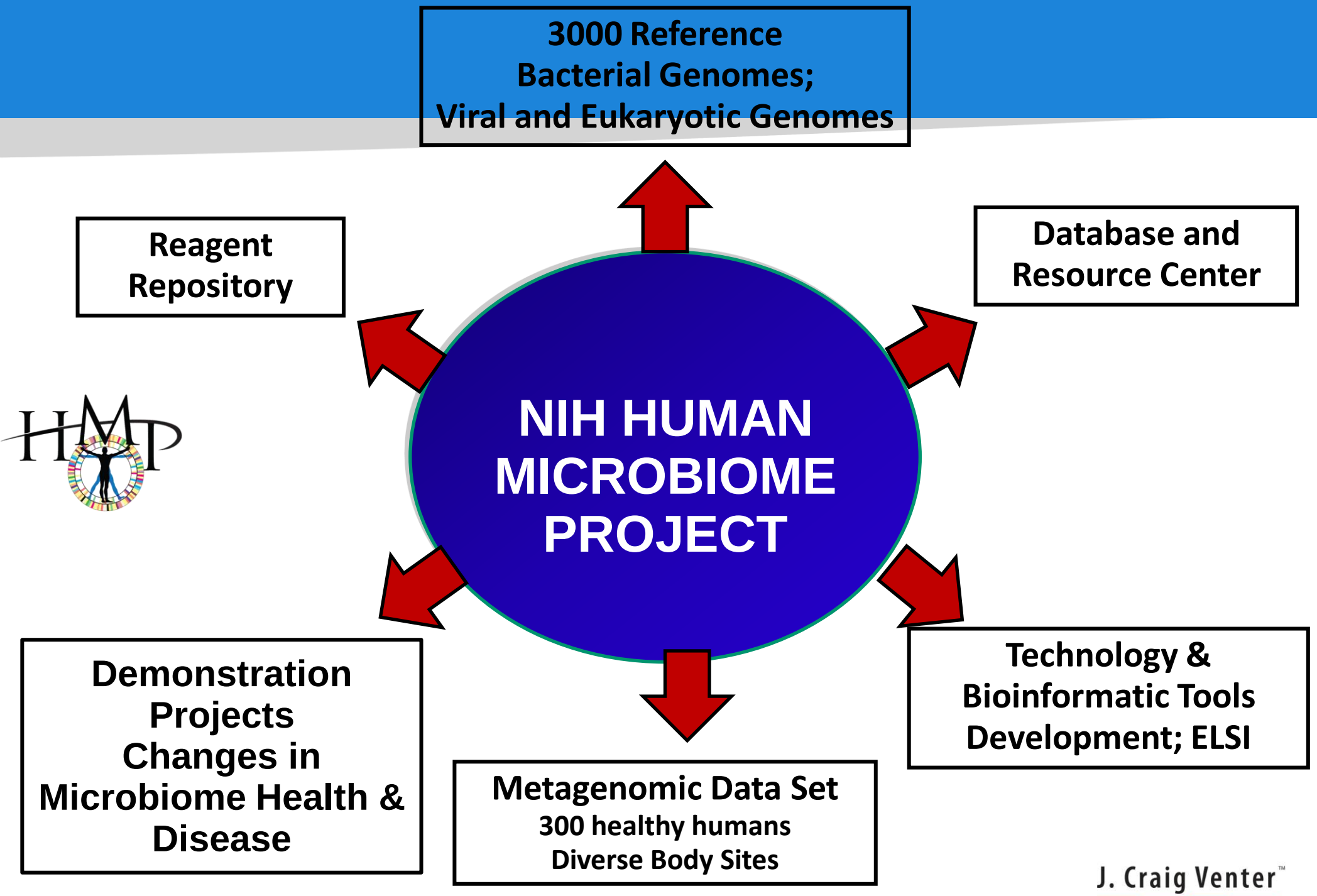
  External Scientific Advisory Group, Workshops.
- **International Consortium**

  Australia, Canada, China, EC, France, Ireland, Japan, Korea, US

- http://nihroadmap.nih.gov/hmp
- http://www.human-microbiome.org/#

J. Craig Venter INSTITUTE

Slide courtesy Maria Giovanni-NIAID



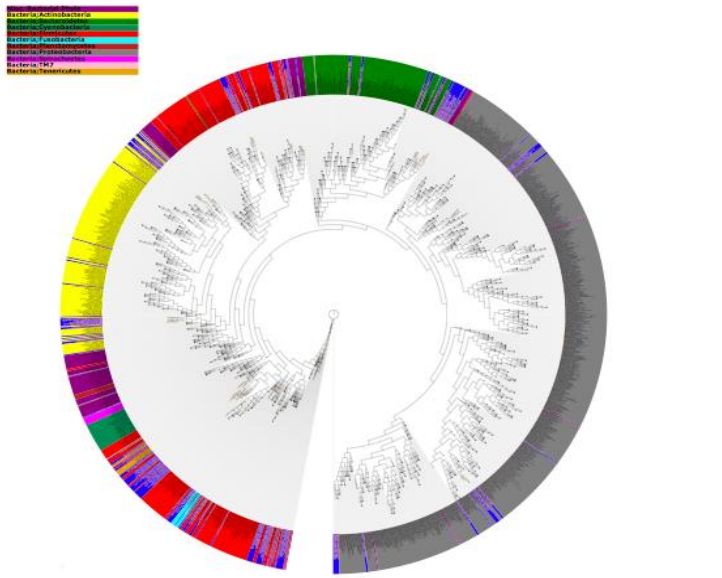
Slide courtesy Maria Giovanni-NIAID

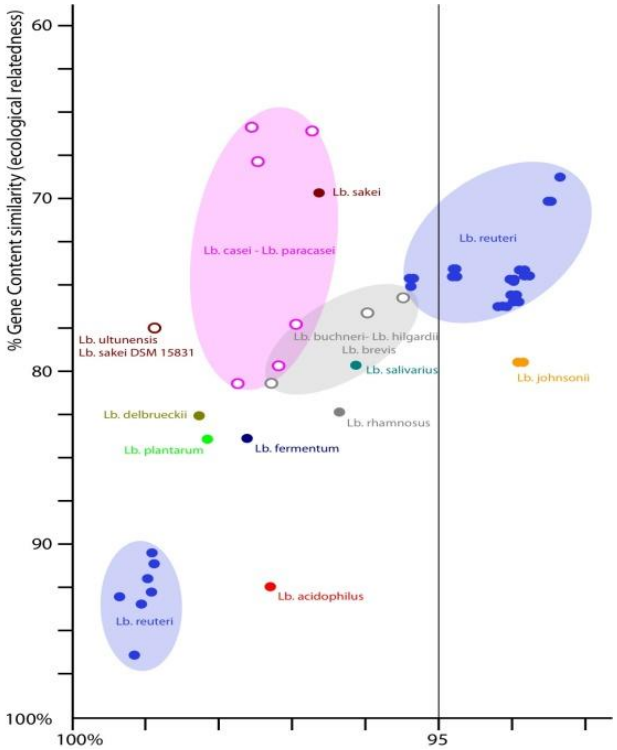
J. Craig Venter™ INSTITUTE

- **Reference Strains:** Generate complete genomes from > 3000 prokaryotes.
- Build our understanding of those recognized through 16S profiles
- Provide for interpretation of metagenomics and other “omics” data
- Sequence reference phage, viruses and eukaryotes

A Catalog of Reference Genomes from the Human Microbiome

178 genomes
~550,000 genes
Nelson et al., Science
May 21, 2010





Human Microbiome Projects Catalog

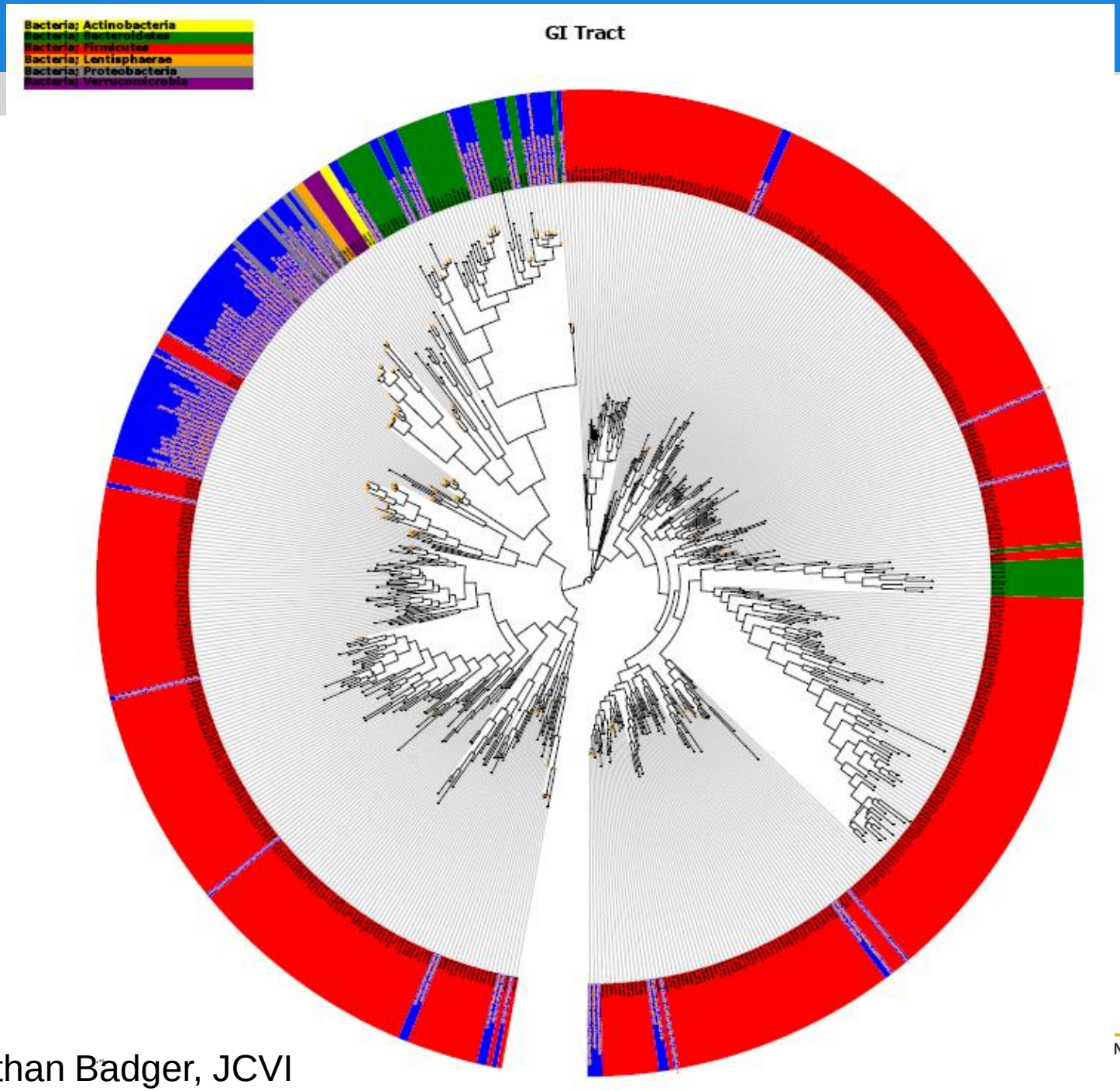
Search Field: is not empty is empty Quick Search

Advanced Search Export to Excel

HMP ID	Organism Name	Body Site	HMP Project Status	Finishing Goal	NCBI Project ID	NCBI Submission Status	Genbank ID	Gene Count	IMG/HMP ID	Sequencing Center	Funding Source	Strain Repository	Cross Ref ID
0001	Abiotrophia defectiva ATCC 49176	Oral	Complete	Level 2: High-Quality Draft	33011	6	ACIN00000000	3346	643886181	Washington Univ, USA	NIH-HMP Jumpstart Supplement	ATCC 49176	HOMD tax_38
0002	Abiotrophia para-adiacens	Airways	Targeted			0				USA	NIH-HMP		
0003	Abiotrophia sp.	Airways	Targeted			0				USA	NIH-HMP		
0004	Achromobacter piechaudii ATCC 43553	Airways	In Progress	Level 2: High-Quality Draft		0				BCM-HGSC, USA	NIH-HMP Jumpstart Supplement	ATCC 43553	
0005	Achromobacter xylosoxidans C54	Airways	In Progress	Level 5: Noncontiguous Finished	38739	2				Broad Institute, USA	NIH-HMP Jumpstart Supplement	BEI Shipping	HOMD tax_34
0006	Achromobacter xylosoxidans	Airways	Targeted			0				USA	NIH-HMP		HOMD tax_34
0007	Achromobacter xylosoxidans	Airways	Targeted			0				USA	NIH-HMP		HOMD tax_34
0008	Acidaminococcus sp. D21	Gastrointestinal tract	Draft	Level 3: Improved-High-Quality Draft	34117	6	ACGB00000000	2055	643886056	Broad Institute, USA	NIH-NHGRI	BEI HM-81	
0009	Acidovorax sp.	Skin	Targeted			0				USA	NIH-HMP		

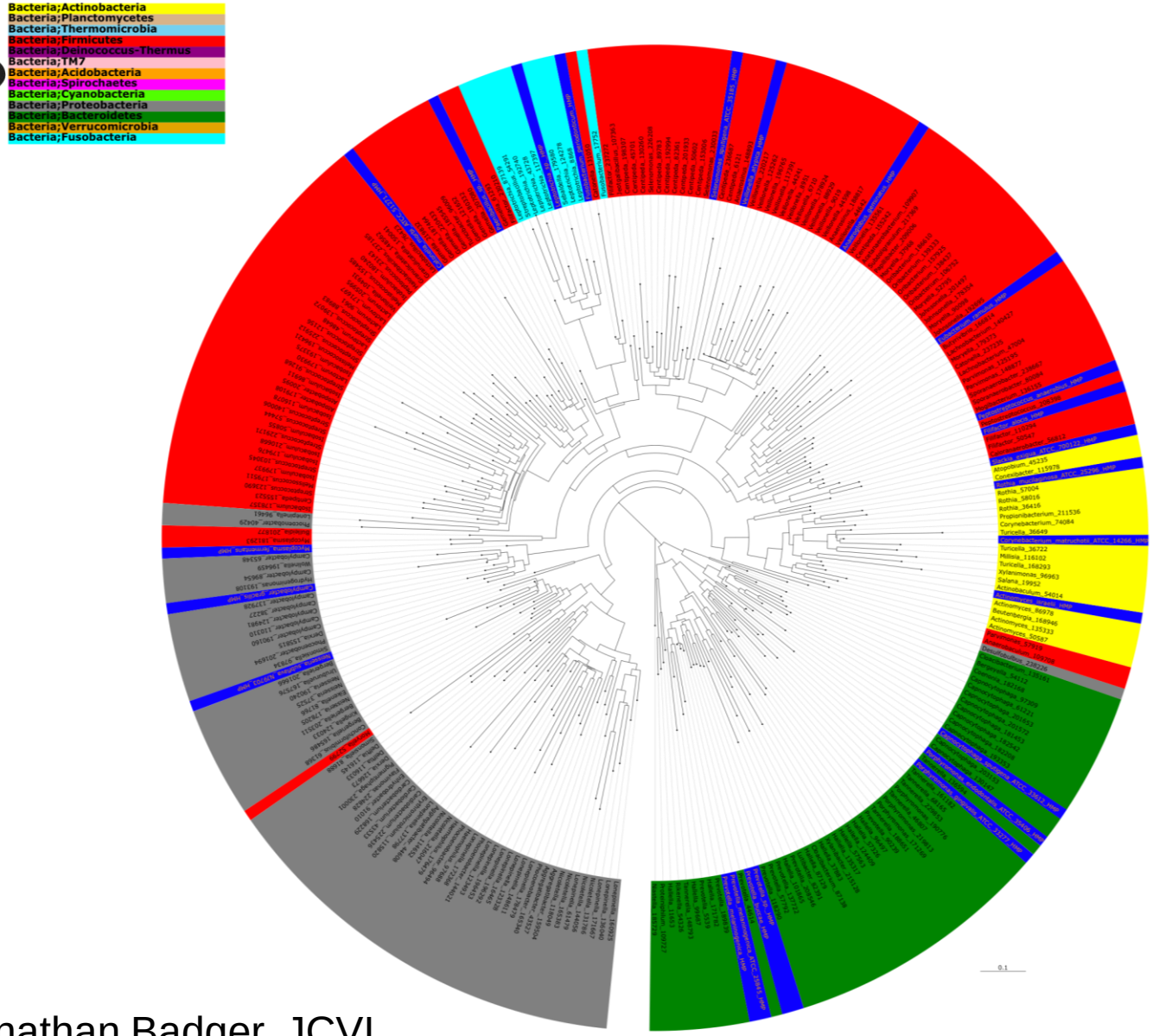
Count: 1295 HMP Master List

Phylogenetic Analysis – GI Tract



Courtesy Jonathan Badger, JCVI

ORAL COMMUNITY SURVEYS

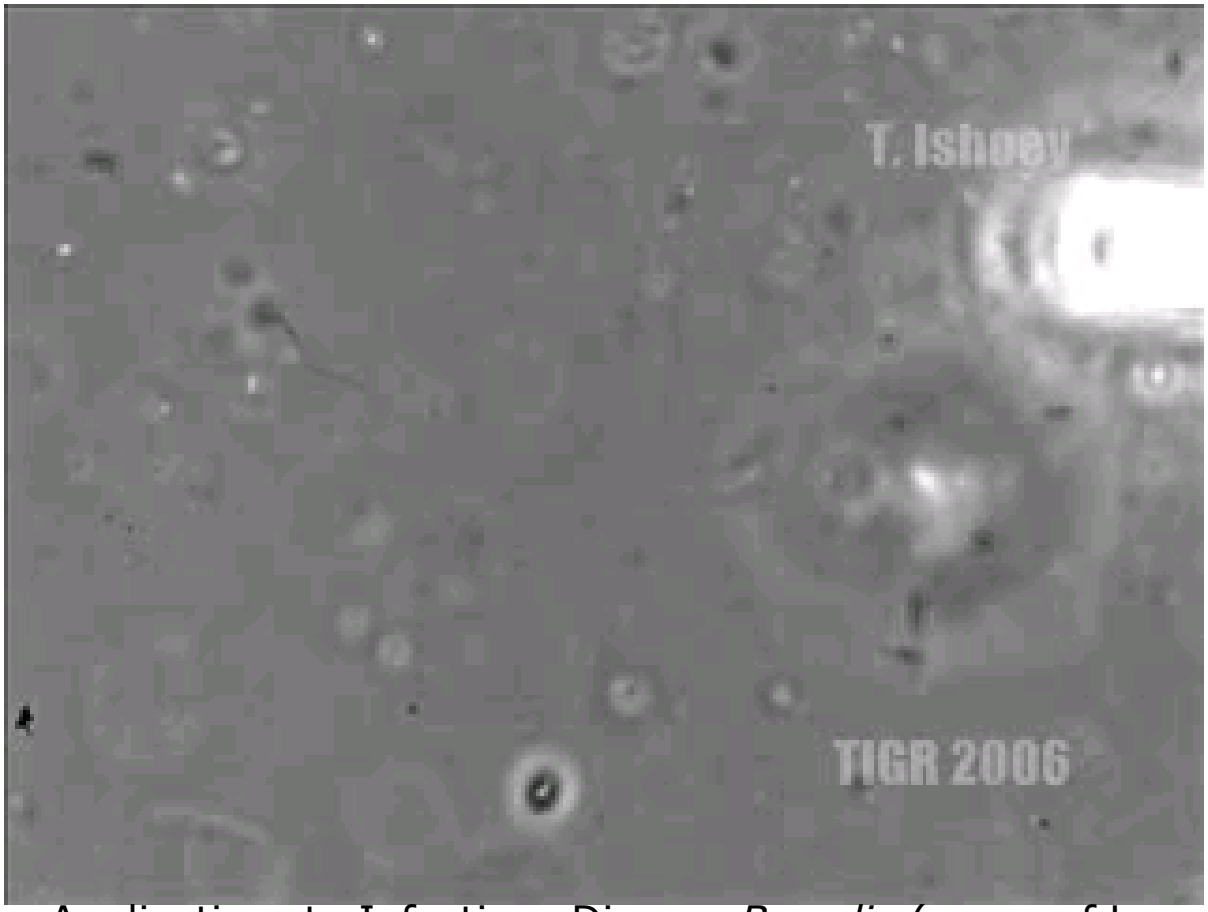


Courtesy Jonathan Badger, JCVI

J. Craig Venter INSTITUTE

Method development for accessing reference genomes

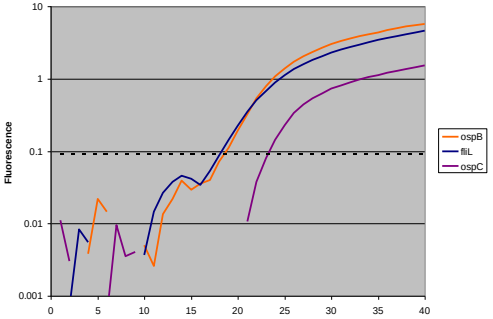
- homogenized tick midgut



Amplify DNA

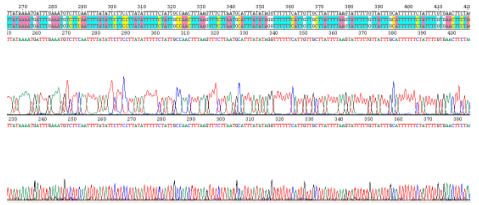
MDA

Genotype by PCR



chromosome and plasmids lp54 and cp26

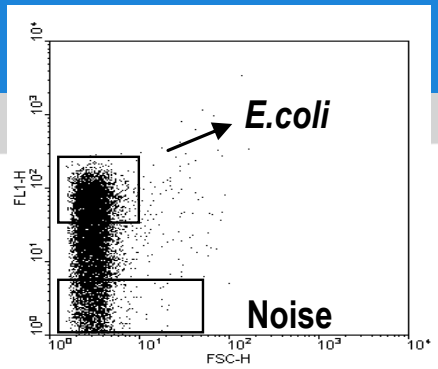
Genomic Sequencing



- Applications to Infectious Disease *Borrelia* (cause of Lyme disease)
  D Qiu, X Yang, B J. Luft, S Schutzer, JJ. Dunn, W Qiu

J. Craig Venter™ INSTITUTE

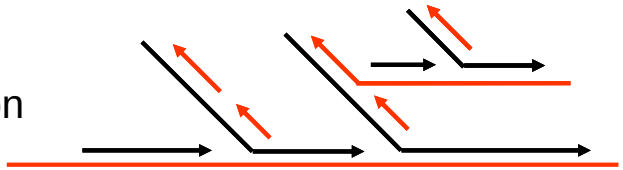
Single Cell Sequencing



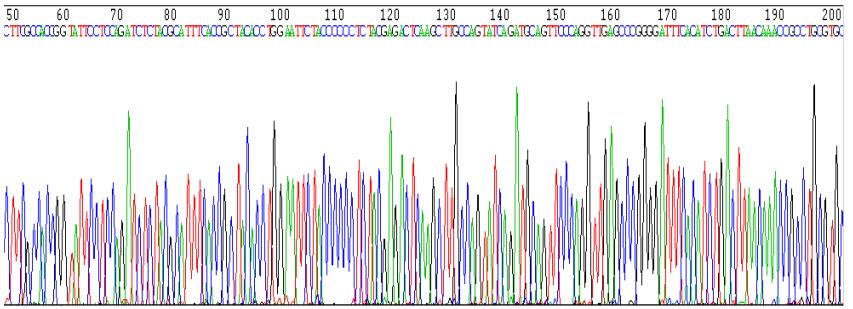
Flow sort single cells

Multiple displacement amplification (MDA)

>10^9 fold amplification



Genotype and Sequence



Raghunathan A, Ferguson, H.R., Bornarth, C.J., Driscoll. M., and Lasken , RS
Applied and Environmental Microbiology (2005) Vol. 71, 3342-3347

J. Craig Venter INSTITUTE

Results for 384 well plate

- **119 single cell amplified bacterial genomes obtained**
- **Confirmation that bacteria were derived from GI tract**

  100/119 have >99% identity to known 16S rRNA gene sequences
  (Eckburg fecal 16S library, Science, 2005)

  Taxonomic diversity observed: Firmicutes (*Clostridium* sp., *Eubacterium* sp., *Lactobacillus* sp.), Bacteroidetes (*B. fragilus* group, Flavobacteriales), Proteobacteria, Verrucomicrobia

- **Confirmation that many genomes are from novel uncultured species**

  21%: >99% identity to sequenced genomes

  57%: 90-99% identity to sequenced genomes

  22%: 80-90% identity to sequenced genomes

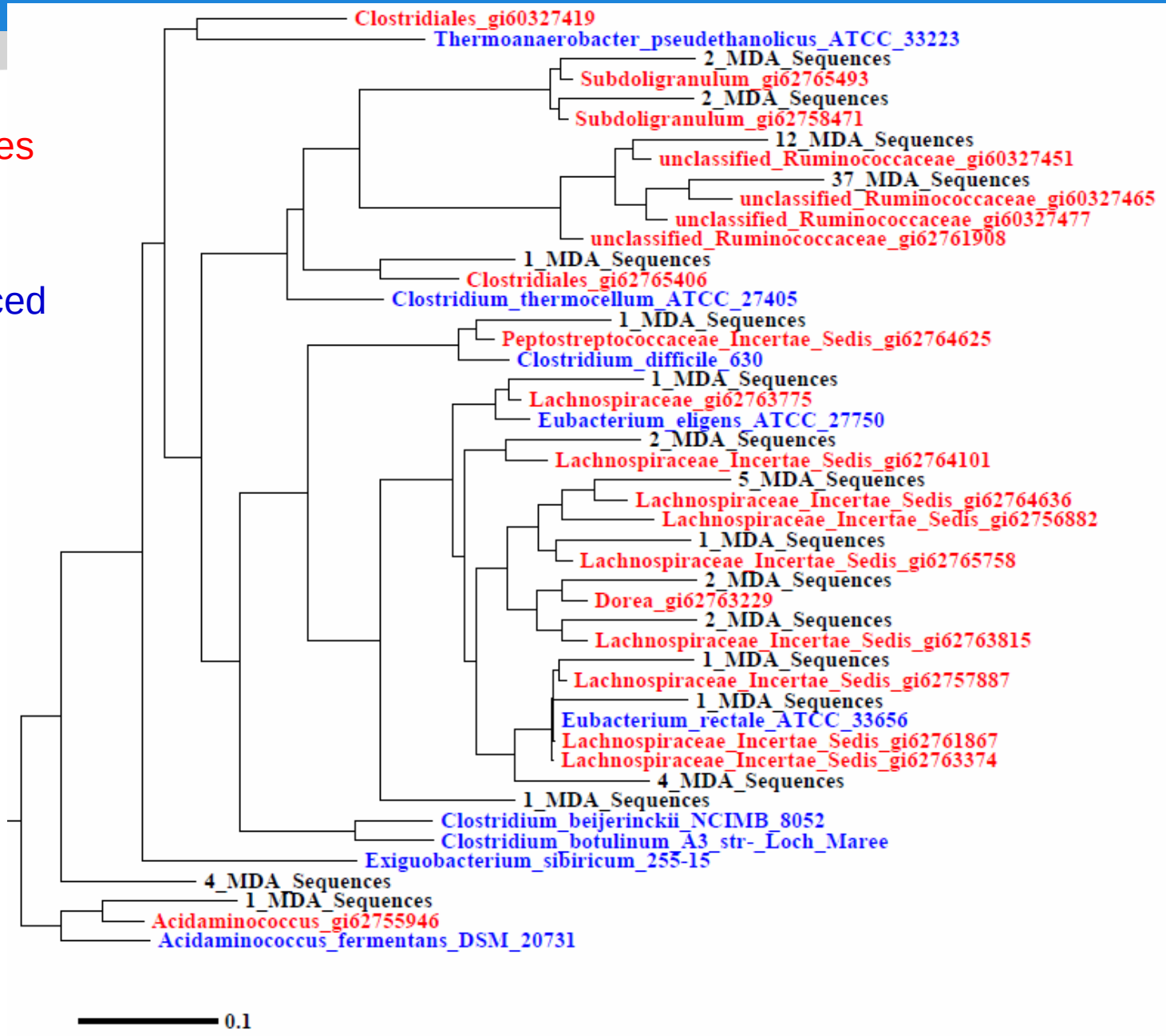
J. Craig Venter™ INSTITUTE

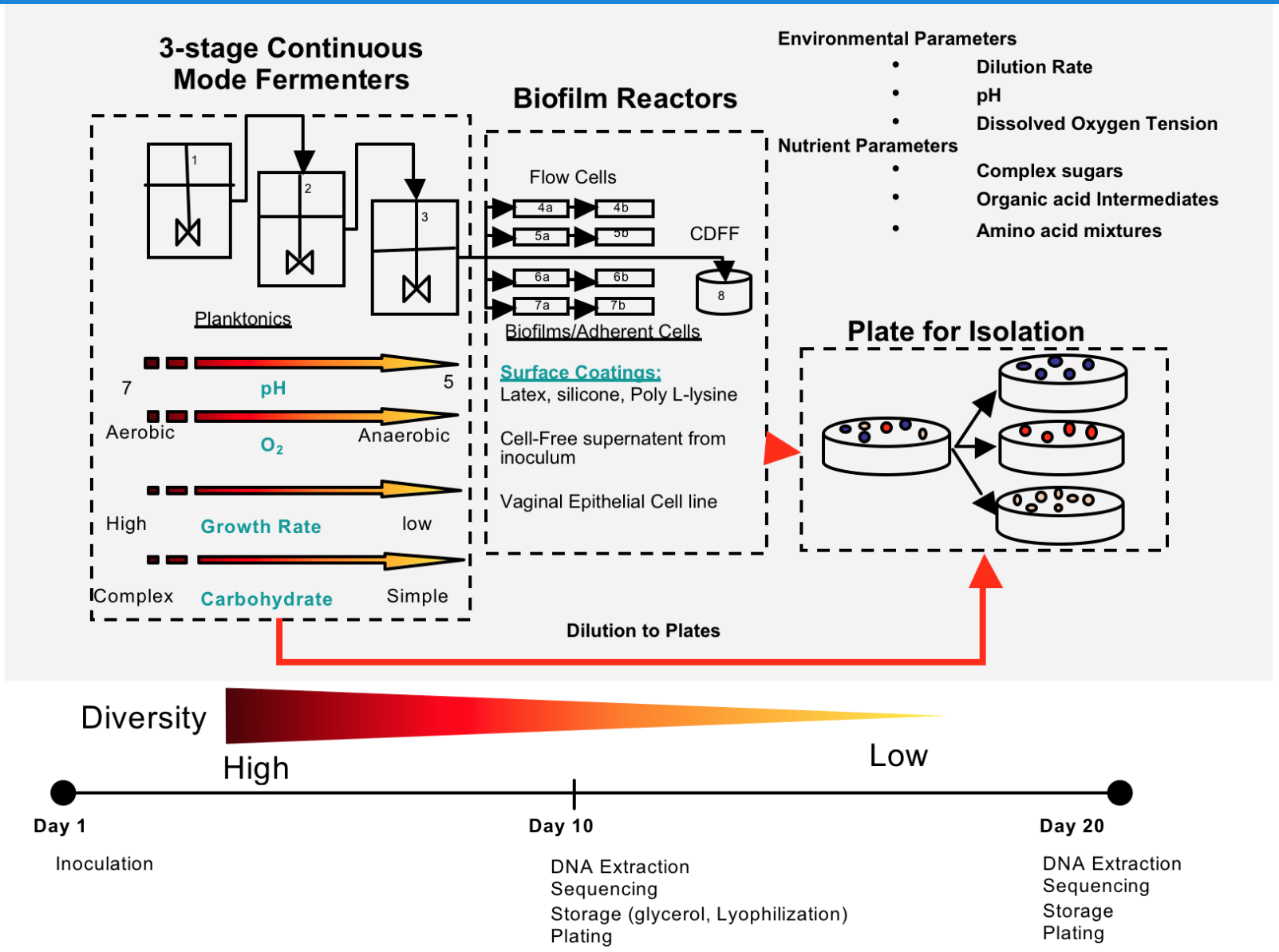
Example of taxonomic analysis

Single cell Clostridiales MDAs fall into several taxonomic groups

Red - 16S fecal sequences (Eckburg)

Blue – 16S from sequenced genomes





By varying conditions in the chemostats, a strong selection will occur, and by feeding this population into different biofilm reactors, one can isolate many different (and new) microbes not seen by more traditional methods.

J. Craig Venter INSTITUTE

Human Health and Disease

- Progression of esophageal cancer
- Bacterial vaginosis and pre-term babies
- Neonatal microbiome/necrotizing enterocolitis
- Nasopharynx microbiome and vaccination in children
- Skin microbiome, acne and psoriasis
- Oral diseases including periodontitis
- Obesity
- Crohn’s and inflammatory bowel disease
- Colon cancer
- Diabetes

J. Craig Venter™ INSTITUTE

Applications for disease conditions

Distal esophagus important anatomic locus where gastric acid reflux causes

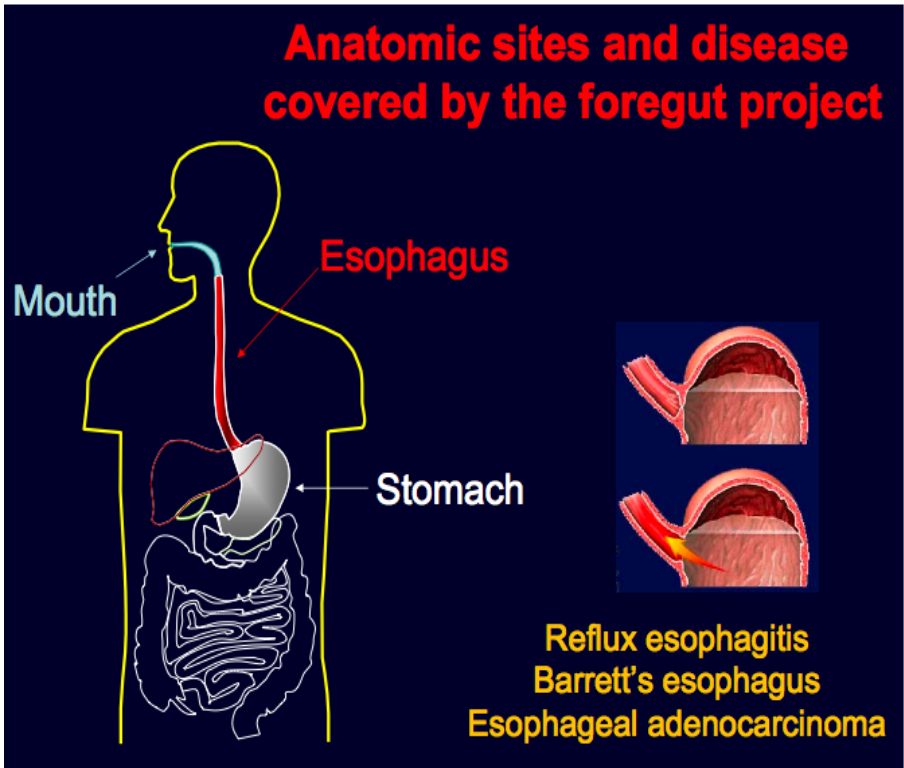
- Reflux esophagitis (RE)
- Barrett's esophagus (BE) and downstream sequelum
- Esophageal adenocarcinoma (EA) fastest rising malignancy

Incidence of EA has increased 6-fold in US since 1970's
Cause remains unknown

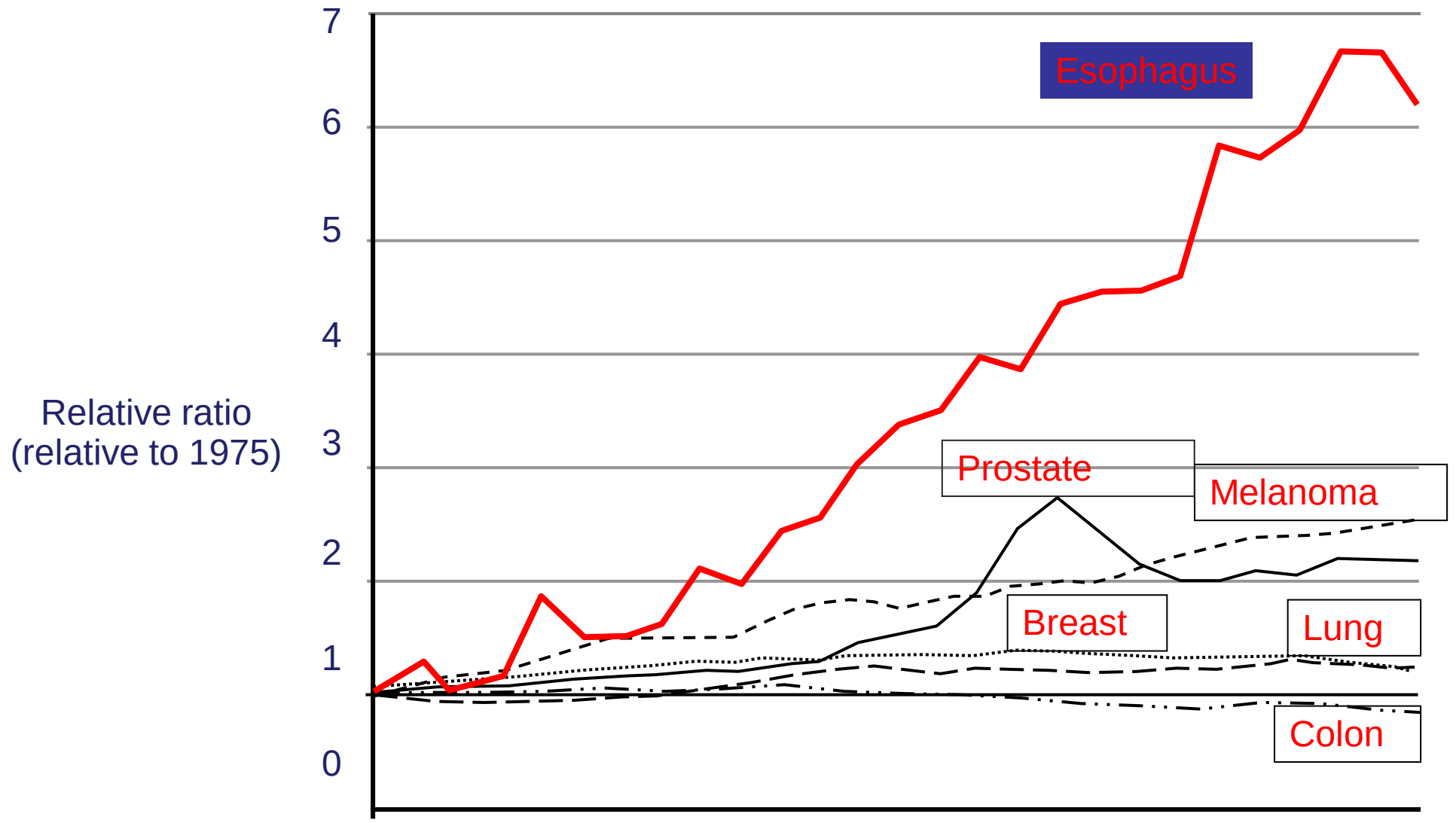
Preliminary studies show two types of microbiotas

Type I normal esophagus

Type II correlated with RE and EA

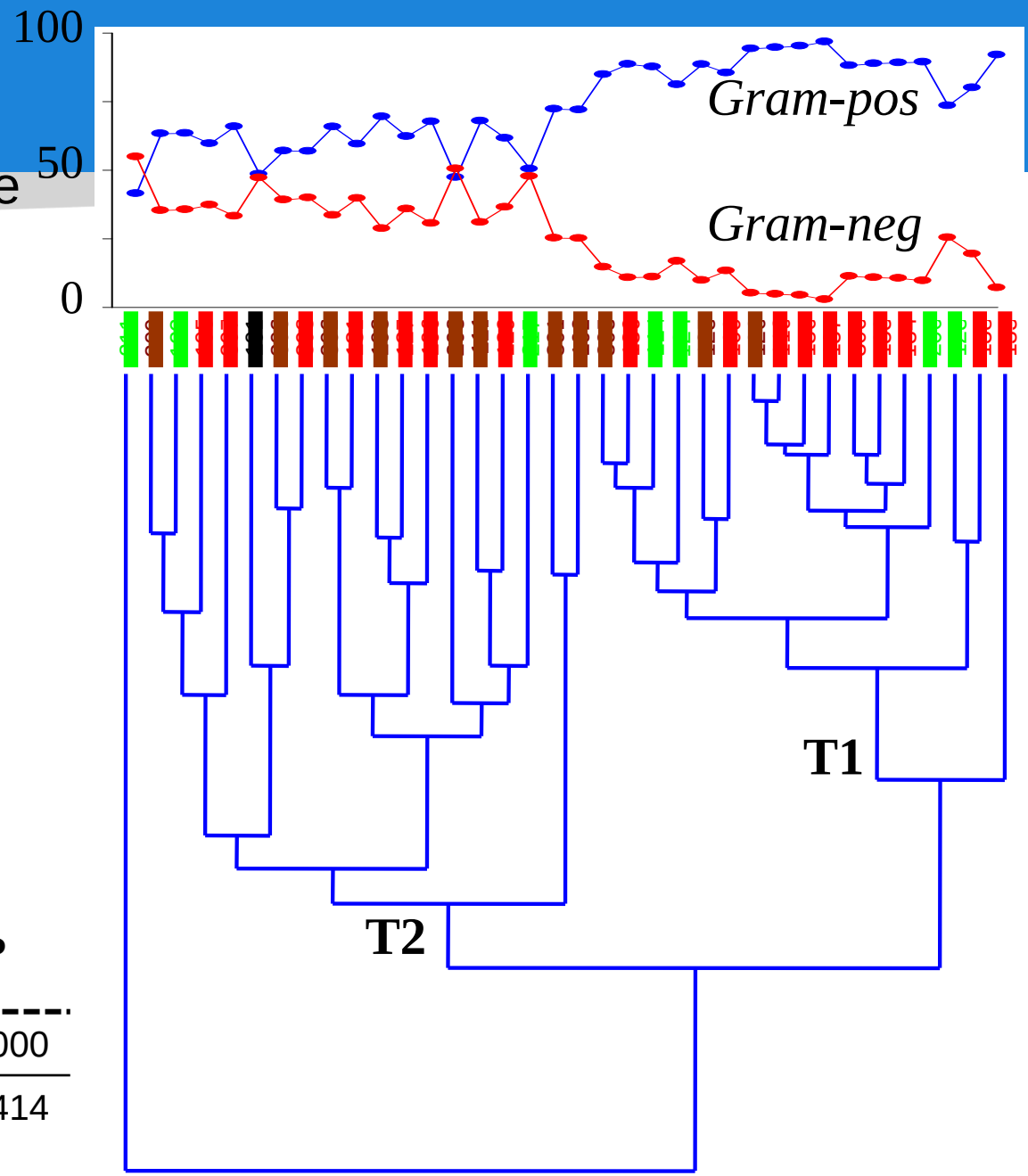


Relative change in incidence of common cancers in USA since 1975



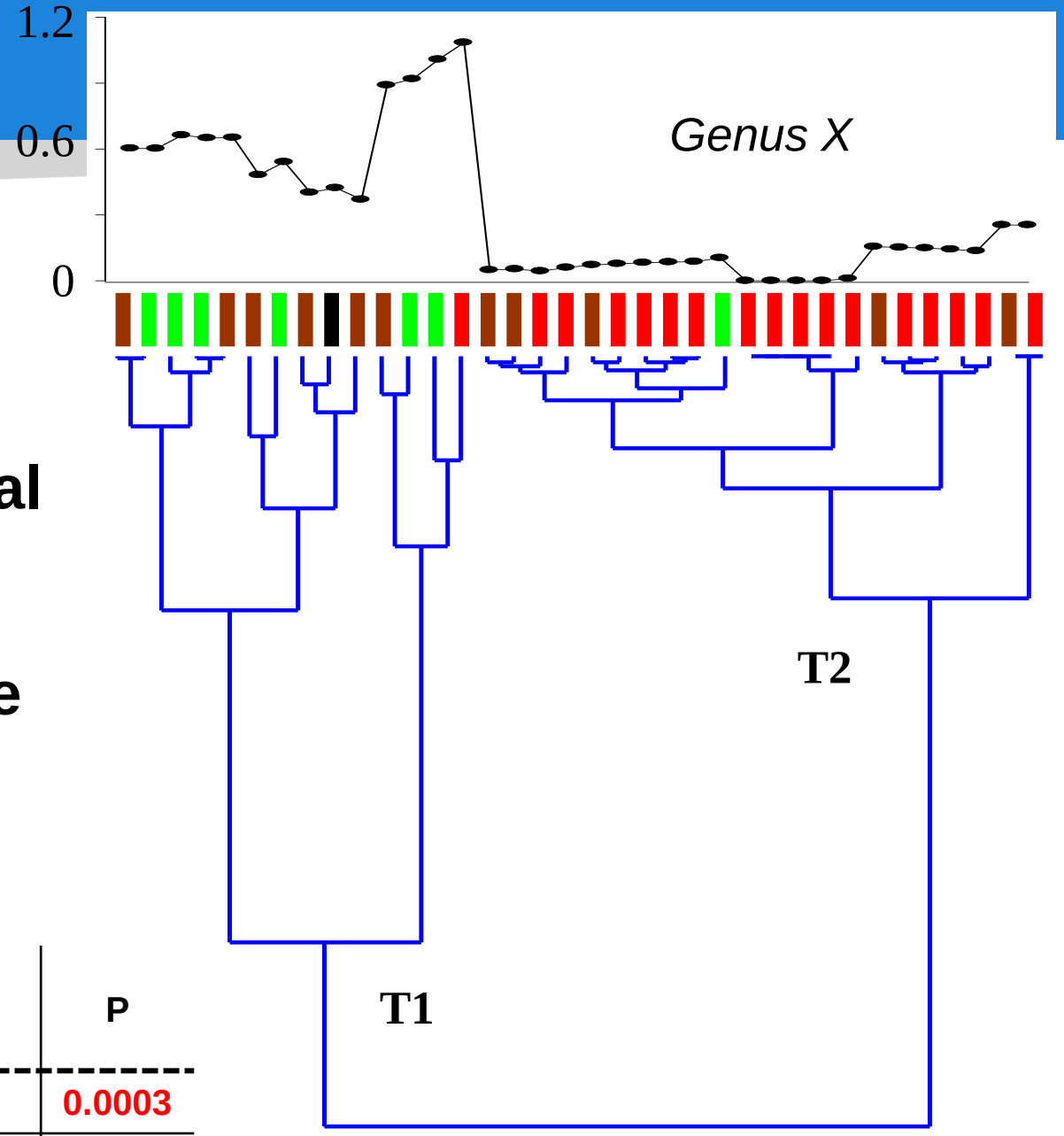
Pohl H. J.Nat.Ca Insti, 97:142-146. 2005

Correlation between microbiome types and phenotypes based on the relative abundance of all major phyla



	T1	T2	OR	P
NL	4 (11)	2 (1)		
RE	10 (5)	7 (7)	1.4	1.0000
BE/EA	3 (4)	9 (9)	6.0	0.1414

J. Craig Venter INSTITUTE



Detection of natural bacterial groups based on the relative abundance of Genus X

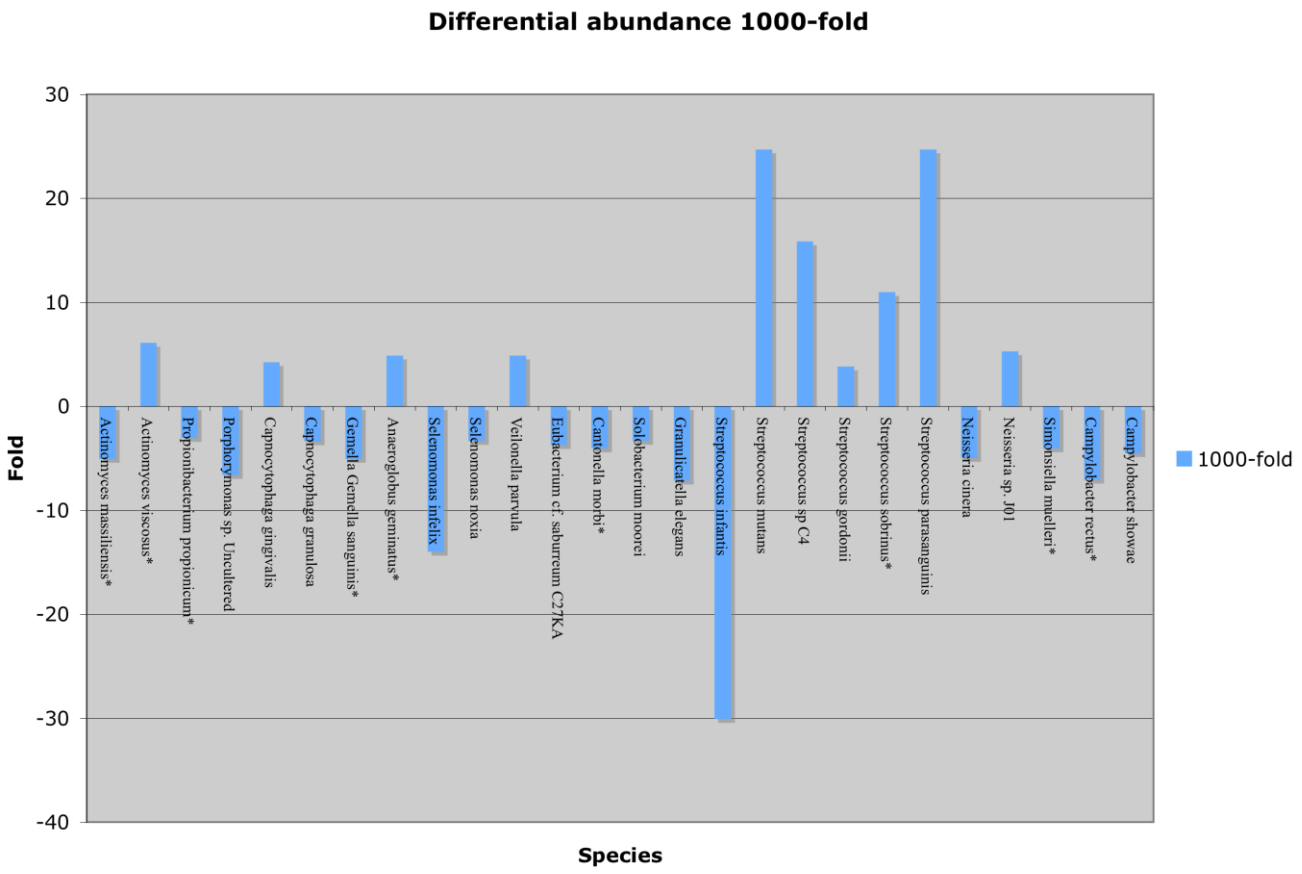
	T1	T2	OR	P
NL	6	1		
RE	1	16	90	0.0003
BE/EA	6	5	5.0	0.3156

J. Craig Venter™ INSTITUTE

Bacterial Vaginosis (BV)
JCVI/University of Illinois

- Common syndrome associated with preterm labor and delivery, pelvic inflammatory disease, and acquisition and transmission of HIV.
- Etiology still poorly understood.
- Traditionally thought to be caused by a single agent.
- BV-related pregnancy complications in the US is nearly $1 billion annually.
- The relationship between 16S rDNA phylotypes and functional capabilities unknown.
- Need for consensus on what constitutes a pathogenic bacterial community.
- Colleagues at The University of Illinois recently studied the site-specific microbial composition within the vaginal ecosystem.
- Revealed that the vaginal microbiota is not homogenous throughout the vaginal tract, but differs significantly within an individual with regard to anatomical site.

Oral Cavity:
Enrichment to identify the minor players



Species

1. Actinomyces naeslundii
2. Capnocytophaga granulose
3. Capnocytophaga sputigena
4. Gemella Gemella sanguinus
5. Anaeroglobus geminatus
6. Megaspaera sp.
7. Selenomonas noxia
8. Veilonella dispar
9. Eubacterium yurii
10. Parvimonas miera
11. Granulicatella elegans
12. Str. Constellatus
13. Str. Infantis
14. Str. Mutans
15. Str. Oligofermentans
16. Str. Oralis
17. Str. Sobrinus
18. Leptotrichia shakii
19. Neisseria bacilliformis
20. Neiserria flava
21. Neiserria sp. J01
22. Campylobacter rectus
23. Campylobacter showae
24. Prevotella oulurum
25. Prevotella salviae
26. Capnocytophaga gingivalis
27. Catonella morbi
28. Catonella sp.
29. Johnsonella sp.
30. Lachnospiraceae bacterium
31. Abiotrophia defectiva
32. Granulicatella para-adjacens
33. Campylobacter concisus
34. Camplyobacter gracilus
35. Actinomyces massiliensis
36. Actinomyces viscosus
37. Propionibacterium prop.
38. Porphymonas sp.
39. Anaeroglobus geminatus
40. Selenomonas infelix
41. Selenomonas noxia
42. Veilonella parvula
43. Eubacterium sabureum
44. Solobacterium moorei
45. Str. Gordonii
46. Str. Parasanguinis
47. Neiserria cinera
48. Simonsiella muelleri

Scott Peterson JCVI

J. Craig Venter INSTITUTE

Summary:
Example of findings from Crohn's

Sokol et al., 2008.
Proc Natl Acad Sci U S A. 2008 Oct 28;105(43):16731-6. Epub 2008 Oct 20.
***Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients.**

Pathogenesis involves ongoing activation of mucosal immune system.

Metagenomics study started in 2005.

Metagenomics, FISH, flow cytometry - revealed decrease in microbial diversity in CD patients; primarily reduction of Firmicutes, in particular the *C. leptum* group.

Recent study proposed countering dysbiosis with commensal *F. prausnitzii*

Challenges with studying the human microbiome

- Study groups, IRB, diversity.
- Removal of Human DNA, personal information….
- Interpretation of data across different groups, worldwide.
- Sample availability and quantity
- Use of Multiple Displacement Amplification (MDA) for small quantities of DNA – bias?
- New sequencing approaches encourage use of variable regions rather than full-length sequences
- Having enough reference genomes for scaffolding.
- Need for large scale culturing efforts based on 16S surveys and metagenomics
- Bioinformatics tool development
- Large scale approaches for screening the functional component

Advantages and disadvantages in metagenomic applications

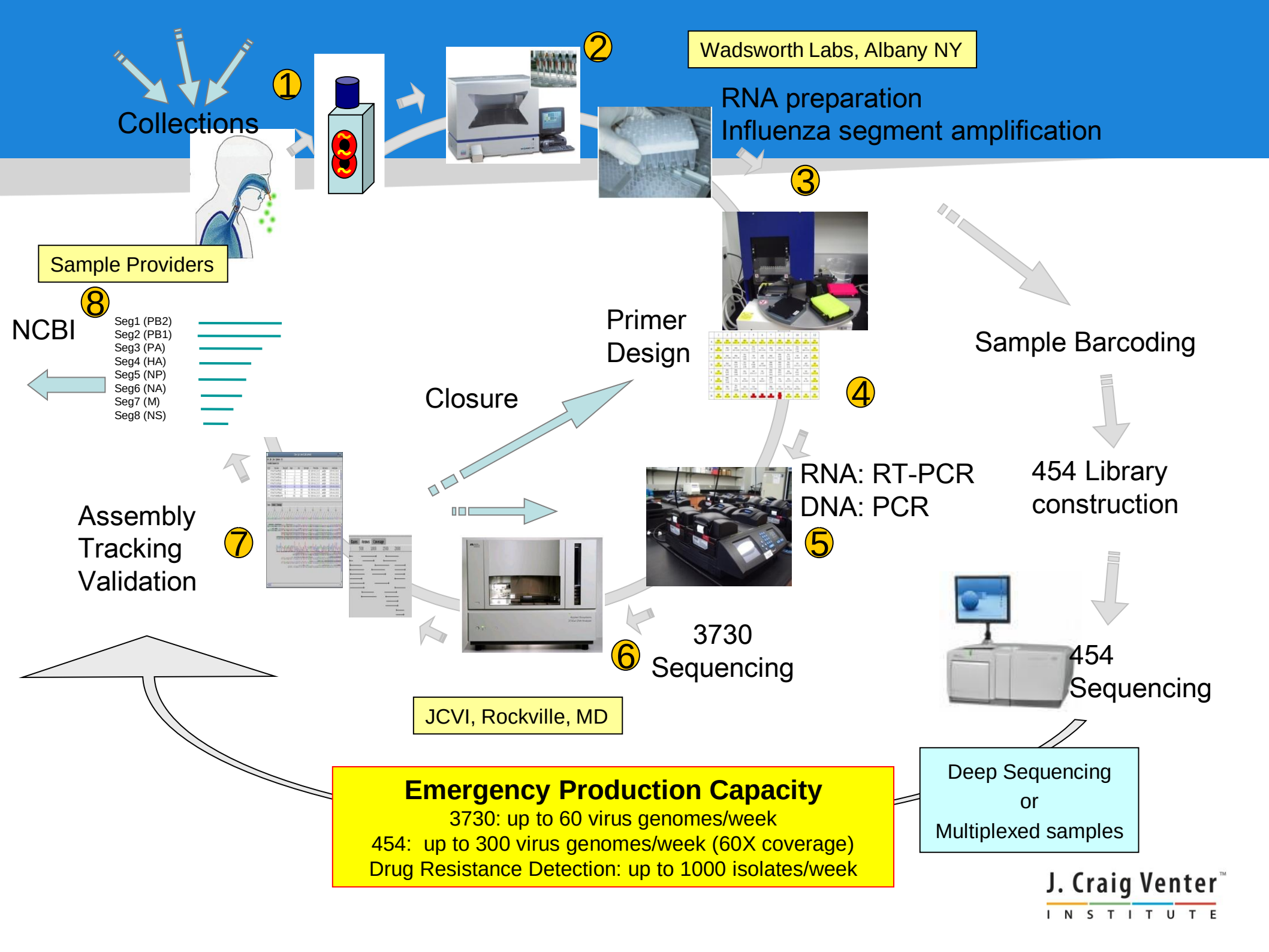
- Greater depth of coverage at a much cheaper cost
- Large amount of sequences that are generated allow for barcoding and multiplex approaches so that multiple samples can be done simultaneously
- Rate of data generation is very quick, versus rate of data analysis
- Assemblers often cannot handle the high depth of sequences
- Not having enough reference genomes for scaffolding to interpret metagenomic data.
- Not closing genomes; how much are we losing in the process?

Other Programs at the JCVI

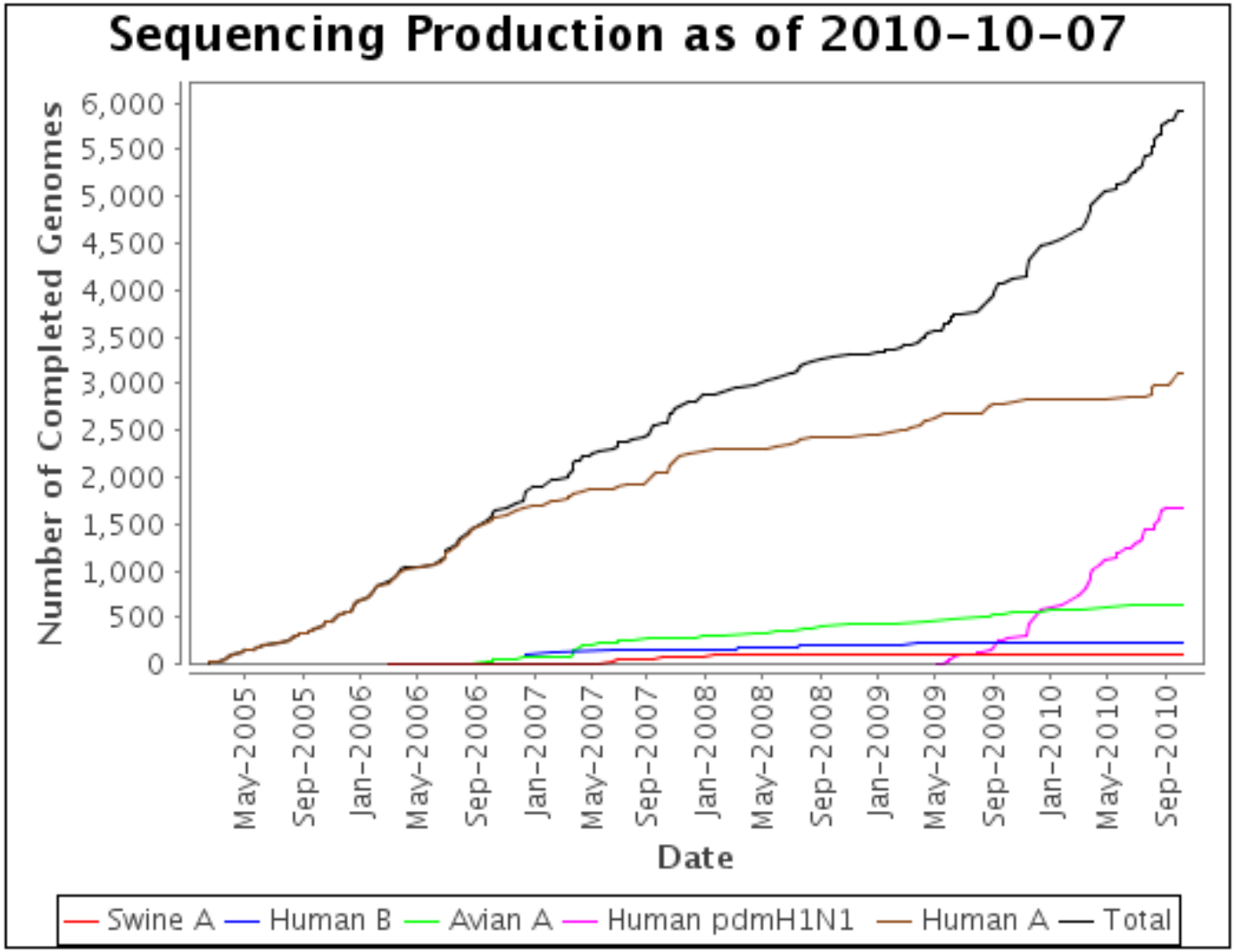
www.niaid.nih.gov/LabsAndResources/resources/gsc/

NIAID Infectious Disease Genomics Program

To provide comprehensive “omics” resources and reagents to the scientific community for basic and applied research in infectious diseases to understand biology of the pathogen, pathogenesis, pathogen-host interaction, and develop potential targets for drugs, vaccines and diagnostics.


Collections
1
2
Wadsworth Labs, Albany NY
RNA preparation
Influenza segment amplification
3
Sample Providers
8
NCBI
Seg1 (PB2)
Seg2 (PB1)
Seg3 (PA)
Seg4 (HA)
Seg5 (NP)
Seg6 (NA)
Seg7 (M)
Seg8 (NS)
Primer
Design
4
Sample Barcoding
Closure
RNA: RT-PCR
DNA: PCR
5
454 Library
construction
Assembly
Tracking
Validation
7
3730
Sequencing
6
454
Sequencing
JCVI, Rockville, MD
Emergency Production Capacity
3730: up to 60 virus genomes/week
454: up to 300 virus genomes/week (60X coverage)
Drug Resistance Detection: up to 1000 isolates/week
Deep Sequencing
or
Multiplexed samples
J. Craig Venter
INSTITUTE

Sequencing of *Influenza* viruses



J. Craig Venter INSTITUTE

JCVI Synthetic Genomics program

Objective: Synthesize a cell with only the machinery necessary for independent life

Why synthesize a minimal cell?

- Knowing what every gene does will allow us to better understand how cells work
- A minimal cell can be used as a launching pad for making more complex and useful organisms
- Producing a minimal cell may help us better understand how to design synthetic cells with useful properties (sequester carbon dioxide, or produce energy, pharmaceuticals and industrial compounds)

J. Craig Venter™ INSTITUTE

Rapid gene synthesis from oligonucleotides

Generating a synthetic genome by whole genome assembly: ϕX174 bacteriophage from synthetic oligonucleotides

Hamilton O. Smith, Clyde A. Hutchison III†, Cynthia Pfannkoch, and J. Craig Venter‡

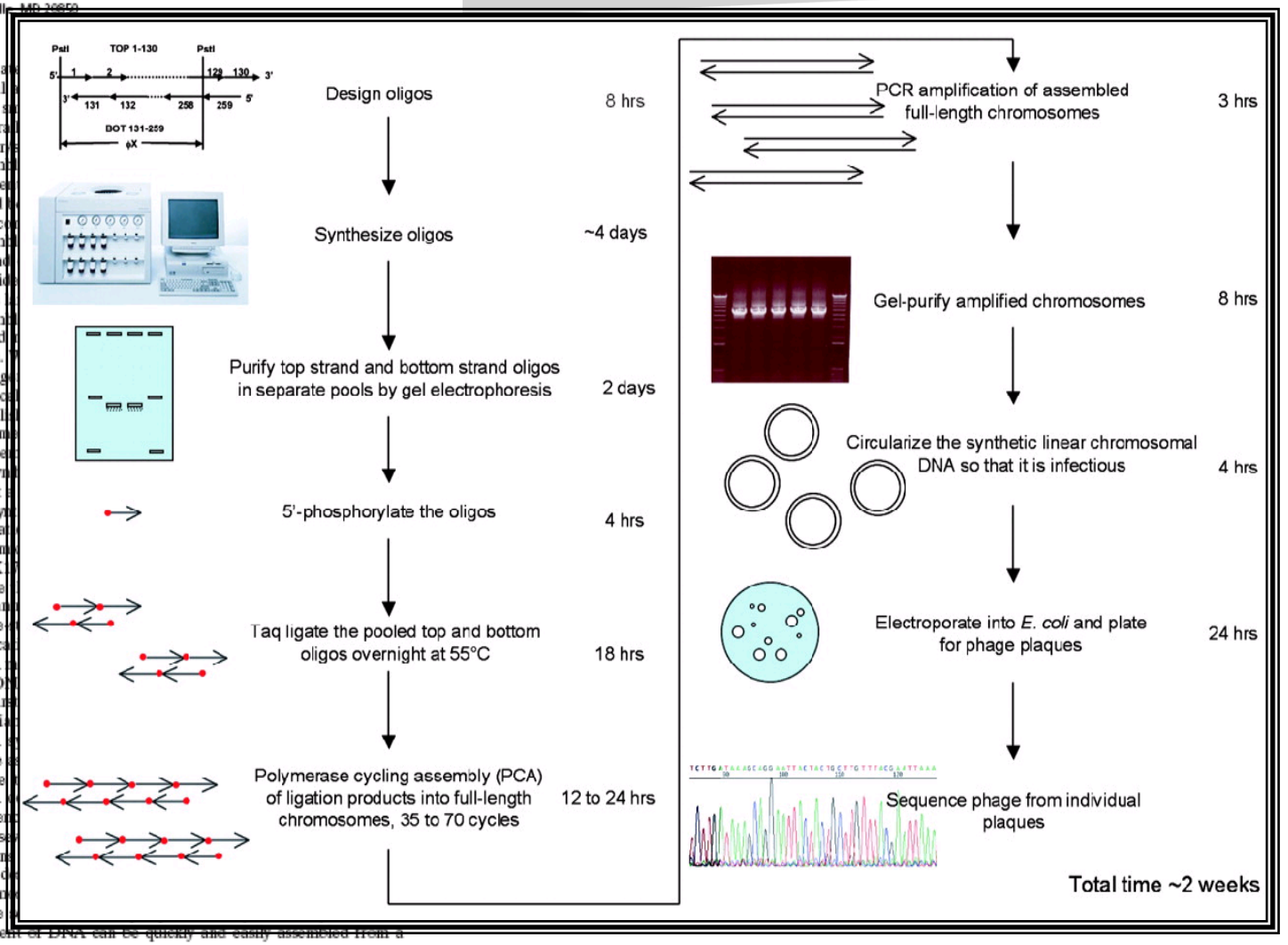
Institute for Biological Energy Alternatives, 1901 Research Boulevard, Suite 600, Rockville, MD 20850

Contributed by J. Craig Venter, November 3, 2003

We have improved upon the methodology and dramatically shortened the time required for accurate assembly of 5- to 6-kb segments of DNA from synthetic oligonucleotides. As a test of this methodology, we have established conditions for the rapid (14-day) assembly of the complete infectious genome of bacteriophage ϕX174 (5,386 bp) from a single pool of chemically synthesized oligonucleotides. The procedure involves three key steps: (*I*) gel purification of pooled oligonucleotides to reduce contamination with molecules of incorrect chain length, (*II*) ligation of the oligonucleotides under stringent annealing conditions (55°C) to select against annealing of molecules with incorrect sequences, and (*III*) assembly of ligation products into full-length genomes by polymerase cycling assembly, a nonexponential reaction in which each terminal oligonucleotide can be extended only once to produce a full-length molecule. We observed a discrete band of full-length assemblies upon gel analysis of the polymerase cycling assembly product, without any PCR amplification. PCR amplification was then used to obtain larger amounts of pure full-length genomes for circularization and infectivity measurements. The synthetic DNA had a lower infectivity than natural DNA, indicating approximately one lethal error per 500 bp. However, fully infectious ϕX174 virions were recovered after electroporation into *Escherichia coli*. Sequence analysis of several infectious isolates verified the accuracy of these synthetic genomes. One such isolate had exactly the intended sequence. We propose to assemble larger genomes by joining separately assembled 5- to 6-kb segments; ≈60 such segments would be required for a minimal cellular genome.

Chemical synthesis of life in the laboratory has been a standing challenge to synthetic organic chemistry since Wöhler's synthesis of urea in 1828 (1), and the doctrine of spontaneous generation was put to rest by an address by Louis Pasteur in 1864.§ With an understanding of the genetic role of DNA, much work has focused on the synthesis of oligonucleotides and genes. The synthesis of the 207-bp gene for tyrosine suppressor tRNA in 1979 by Khorana and 17 coworkers (2) was a monumental undertaking. Since then, the automated DNA synthesizer has been developed based on fundamental advances in synthetic methods from the laboratories of Letsinger (3, 4) and Caruthers (5, 6).

In 1999 we described a minimal prokaryotic genome based on results from random whole genome transposon mutagenesis that inactivated one gene per cell (7). By using this approach, ≈300 essential genes for self-replicating cellular life were described, and we proposed to make a synthetic chromosome to test the viability of this hypothesis (7). Before attempting synthesis of a microbial chromosome, we commissioned an independent bioethical review of our proposed scientific plan (8). After >1 year of deliberation, the reviewers concluded that we were taking a reasonable scientific approach to an important biological question. The broader implications of the creation of life in the laboratory can now be considered a realistic possibility. However, there are several technical barriers to the synthesis of microbial chromosome-sized stretches of DNA that are hundreds of thousands to millions of nucleotides long, the most notable being the contamination of the oligonucleotides by



Abbreviations: RF, replicative form of DNA; PCA, polymerase cycling assembly; synϕX, synthetic ϕX174.

†Present address: Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC 27599-7290.

‡To whom correspondence should be addressed. E-mail: jcventer@tcag.org.

§Pasteur, L., Sorbonne Scientific Soiree, Apr. 7, 1864, Paris.

www.pnas.org/cgi/doi/10.1073/pnas.2237126100

Hamilton Smith, Clyde Hutchison, Cindi Pfannkoch, Craig Venter

J. Craig Venter™ INSTITUTE

RESEARCH ARTICLE
3 AUGUST 2007 VOL 317 SCIENCE www.sciencemag.org

Genome Transplantation in Bacteria: Changing One Species to Another

29 Feb 2008

Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Clyde A. Hutchison

We have synthesize
named *M. genitaliu*
MG408, which was
selection. To identif

25 Sep 2009

Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast

Carole Lartigue,[1] Sanjay Vashee,[1]† Mikkel A. Algire,[1] Ray-Yuan Chuang,[1] Gwynedd A. Benders,[2] Li Ma,[1] Vladimir N. Noskov,[1] Evgeniya A. Denisova,[1] Daniel G. Gibson,[1] Nacyra Assad-Garcia,[1] Nina Alperovich,[1] David W. Thomas,[1]* Chuck Merryman,[1] Hamilton O. Smith,[2] J. Craig Venter,[1,2] John I. Glass[1]

reliable methods for the assembly and cloning of
much larger synthetic DNA molecules.
Strategy for synthesis and assembly. The na
tive 580,076-bp *M. genitalium* genome sequen
(*Mycoplasma genitalium* G37 ATCC 335
genomic sequence; accession no. L43967)
was partitioned into 101 cassettes of appr
mately 5 to 7 kb in length (Fig. 1) that
individually synthesized, verified by sequer
and then joined together in stages. In g

omously repl
propagation
mid (YCp)
one clone
vector int
grew rob
M. capr
into yea
YCpM
of the
(i) th
M.
gen
pla
ge

assembly, and cloning of a bacterial genome in yeast.
ferred from yeast to a receptive cytoplasm.
plasma mycoides genome as a

I U T E

Sciencexpress Research Article

Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,[1] John I. Glass,[1] Carole Lartigue,[1] Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Mikkel A. Algire,[1] Gwynedd A. Benders,[2] Michael G. Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] Chuck Merryman,[1] Sanjay Vashee,[1] Radha Krishnakumar,[1] Nacyra Assad-Garcia,[1] Cynthia Andrews-Pfannkoch,[1] Evgeniya A. Denisova,[1] Lei Young,[1] Zhi-Qing Qi,[1] Thomas H. Segall-Shapiro,[1] Christopher H. Calvey,[1] Prashanth P. Parmar,[1] Clyde A. Hutchison III,[2] Hamilton O. Smith,[2] J. Craig Venter[1,2*]

[1]The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA. [2]The J. Craig Venter Institute, 10355 Science Center Drive, San Diego, CA 92121, USA.

[*]To whom correspondence should be addressed. E-mail: jcventer@jcvi.org

May 2010

What we have learned during our quest to synthesize a bacterial cell…

J. Craig Venter INSTITUTE

- It will soon be possible to synthesize any sequence you can specify and install it in a cell where it can be expressed.
- Genome construction is NOT going to be the rate limiting step in producing designer cells.
- Work on computational tools for genome and pathway design is needed.

Merging synthetic approaches with other research areas

Current Annual Vaccine Production Cycle

Step 1: Global Surveillance

WHO reference labs characterize thousands of samples using serological and molecular techniques

Year-round

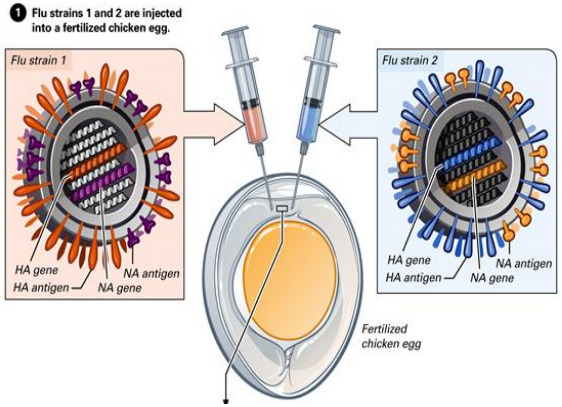
Step 2: Strain Selection

WHO and CDC recommend dominant influenza strains and submit to FDA. FDA selects strains and sends them to vaccine manufacturers.

January-March

Step 3: Egg Adaptation

Vaccine strains are injected into chicken eggs along with standard vaccine backbone strain in order to select high yield reassortants which grow efficiently in eggs



~One month

J. Craig Venter INSTITUTE

Synthetic Genomics Vaccine Production

World Health Organization

National Institute of Allergy and Infectious Diseases
National Institutes of Health

J. Craig Venter INSTITUTE

CDC

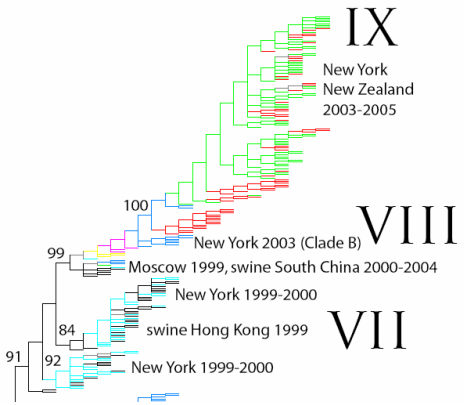
NCBI

Step 1: Global Influenza Sequencing

Labs across the globe provide sequence data about current circulating influenza strains.

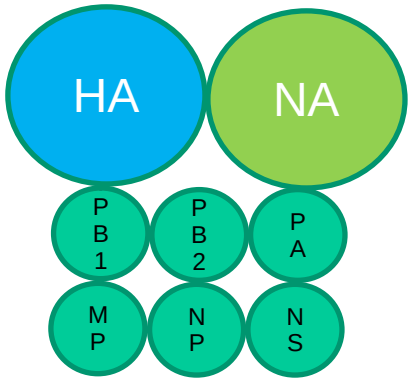
Step 2: Advance Prediction of Dominant Strains

leverage all available sequence information and use bioinformatic tools to characterize and predict dominant strains



Step 3: Synthetic Genomics

Use of synthetic genomics to construct expression plasmids containing the genetic information of predicted dominant strains.



Year-round:

1. Surveillance
2. Advance Prediction,
3. Synthetic Genomics Construction

will make production - ready vaccine seeds available to manufacturers **the day that the WHO recommendations are released**.

This would save at least a month as compared to the current process.

J. Craig Venter™ INSTITUTE

- The intent of this Symposium is to bring together microbial ecologists, population biologists, microbial geneticists, conservation biologists, regulators, and sociologists, as well as scientists directly involved in developing microbial control agents, to better understand the “state of the science” in microbe-based biological control, provide a balanced discussion regarding perceptions of risk, and identify strategies to improve public trust and support for microbial biological control.

J. Craig Venter™
INSTITUTE

Acknowledgments

JCVI TEAM

Barb Methe
Diana Radune
Les Foster
Granger Sutton
Eli Venter
Derrick Fouts
Manolito Torralba
Doug Rusch
Nadia Federova
Kelvin LI
Yu-Hui Rogers
Indresh Singh
Roger Lasken
Scott Peterson
Tanja Davidsen
Jamison McCorrison
Marcus Gillis
Diana Radune
Tim Stockwell
Craig Venter
Shibu Yooseph
Ishwar C.
Jessica Hostetler
Scott Durkin
Ramana Madupu
William Nierman

FUNDING- NIH

GSC award funded in whole or part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900007C.

HMP project supported by Award Number U54 AI-084844-01 administered by the National Institute of Allergy and Infectious Diseases on behalf of the NIH Roadmap Human Microbiome Project.

J. Craig Venter™
INSTITUTE